

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2000, 23:12:37 ; Search time 2502.06 Seconds
(without alignments)
1991.508 Million cell updates/sec

Title: US-08-753-750A-1
Perfect score: 2792
Sequence: 1 atgataatgaatatcatca.....cattgaaatgaagttag 2792

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_bal:*
- 2: gb_bal:*
- 3: gb_bal:*
- 4: gb_bal:*
- 5: gb_bal:*
- 6: gb_bal:*
- 7: gb_bal:*
- 8: gb_bal:*
- 9: gb_bal:*
- 10: gb_bal:*
- 11: gb_bal:*
- 12: gb_bal:*
- 13: gb_bal:*
- 14: gb_bal:*
- 15: gb_bal:*
- 16: gb_bal:*
- 17: gb_bal:*
- 18: gb_bal:*
- 19: gb_bal:*
- 20: gb_bal:*
- 21: gb_bal:*
- 22: gb_bal:*
- 23: gb_bal:*
- 24: gb_bal:*
- 25: gb_bal:*
- 26: gb_bal:*
- 27: gb_bal:*
- 28: gb_bal:*
- 29: gb_bal:*
- 30: gb_bal:*
- 31: gb_bal:*
- 32: gb_bal:*
- 33: gb_bal:*
- 34: gb_bal:*
- 35: gb_bal:*
- 36: gb_bal:*
- 37: gb_bal:*
- 38: gb_bal:*
- 39: gb_bal:*
- 40: gb_bal:*
- 41: gb_bal:*
- 42: gb_bal:*
- 43: gb_bal:*

- 44: gb_htg7:*
- 45: em_htg1:*
- 46: em_htg2:*
- 47: em_htg3:*
- 48: em_hum5:*
- 49: gb_pl3:*
- 50: gb_pr5:*
- 51: gb_htg8:*
- 52: gb_htg9:*
- 53: gb_htg10:*
- 54: gb_htg11:*
- 55: gb_htg12:*
- 56: gb_htg13:*
- 57: gb_htg14:*
- 58: gb_in3:*
- 59: gb_htg15:*
- 60: gb_htg16:*
- 61: gb_htg17:*
- 62: em_htg4:*
- 63: em_htg5:*
- 64: em_htg6:*
- 65: em_htg7:*
- 66: em_hum6:*
- 67: gb_htg18:*
- 68: gb_htg19:*
- 69: gb_htg20:*
- 70: gb_htg21:*
- 71: gb_htg22:*
- 72: gb_htg23:*
- 73: gb_htg24:*
- 74: gb_htg25:*
- 75: gb_htg26:*
- 76: gb_htg27:*
- 77: gb_htg28:*
- 78: gb_htg29:*
- 79: gb_htg30:*
- 80: gb_htg31:*
- 81: gb_vil:*
- 82: gb_v12:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2781	99.6	6378	2 PHU73302	U73302 Pasteurella
2	531	19.0	3541	2 APU16019	U16019 Actinobacil
3	529.4	19.0	2826	5 A56682	A56682 Sequence 1
4	529.4	19.0	2833	1 APTBP1	Z49708 A.pleuropne
5	452.2	16.2	3447	1 APDNATFEB	Z54191 A.pleuropne
6	450.6	16.1	5083	2 APU16017	U16017 Actinobacil
7	411	14.7	5200	2 APL17915	Y17915 Actinobacil
8	345.2	12.4	10205	2 U32780	U32780 Haemophilus
9	340.8	12.2	5099	2 HIU15053	U15053 Haemophilus
10	340.8	12.2	5099	5 I79990	I79990 Sequence 4
11	332.6	11.9	5144	2 HIU15058	U15058 Haemophilus
12	332.6	11.9	5144	5 I79999	I79999 Sequence 10
13	332.4	11.9	4927	1 HIU10882	U10882 Haemophilus
14	331.2	11.9	5009	2 HIU15052	U15052 Haemophilus
15	331.2	11.9	5033	2 HIU15051	U15051 Haemophilus
16	331.2	11.9	5033	5 I79988	I79988 Sequence 2
17	331.2	11.9	5033	5 I79987	I79987 Sequence 1
18	314.4	11.3	4699	5 NGU16260	U16260 Neisseria 9
19	271.8	9.7	3300	2 AE002504	AE002504 Neisseria
20	265.4	9.5	9955	2 NMLBPAG	X79838 N.meningit
21	262.4	9.4	3398	1 NMIROA	X69214 N.meningit
22	261.4	9.4	3171	1 AF049349	AF049349 Neisseria
23	261.4	9.4	5691	2 NMA622491	AL162757 Neisseria
24	258.4	9.3	326301	2	

```

25 235.4 8.4 268 2 AF031694
26 231.4 8.3 2693 1 ACNTEBAA
27 231.4 8.3 2696 5 AR037085
28 231.4 8.3 2696 5 I11990
29 231.4 8.3 2696 5 I21326
30 215.6 7.7 8253 2 AF039315
31 215.6 7.7 8266 2 AF039312
32 193.8 6.9 2800 5 A36256
33 193.8 6.9 4934 1 NMTBP12A
34 193.8 6.9 6114 2 AF058688
35 190.2 6.8 8946 2 AF043131
36 185.4 6.6 8937 2 AF043132
37 181.2 6.5 3537 5 AR072081
38 166.2 6.0 29831 2 AF058689
39 155 5.6 2809 5 A36258
40 155 5.6 5472 1 NMTBP12B
41 150.4 5.4 11109 2 AE020402
42 150.2 5.4 2762 1 NMTBP1B
43 143.8 5.2 2916 1 NG07BPA
44 143.8 5.2 3286 5 AR072080
45 134.2 4.8 2653 1 NMTBPAG

```

ALIGNMENTS

```

RESULT 1
PHU73302 PHU73302 6378 bp DNA BCT 30-DEC-1997
LOCUS Pasteurella haemolytica KNAse T, transferrin binding protein B,
DEFINITION transferrin binding protein A and FIS genes, complete cds.
ACCESSION U73302
VERSION U73302.1 GI:2731430
KEYWORDS
SOURCE Pasteurella haemolytica.
ORGANISM Pasteurella haemolytica.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Pasteurella.
REFERENCE 1 (bases 1 to 6378)
AUTHORS Schryvers,A.B., Ogunnariwo,J.A., Gonzalez,G.C., Woo,T.K.W. and
Lo,R.Y.C.
TITLE Characterization of the Pasteurella haemolytica transferrin
JOURNAL receptor protein genes and the recombinant receptor proteins
REFERENCE 2 (bases 1 to 6378)
AUTHORS Schryvers,A.B., Ogunnariwo,J.A., Woo,T.K.W. and Lo,R.Y.C.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1996) Microbiology and Infectious Diseases,
University of Calgary, 3330 Hospital Drive NW, Calgary, Alberta T2N
4N1, Canada

```

```

FEATURES
    source      Location/Qualifiers
                1..6378
                /organism="Pasteurella haemolytica"
                /strain="h196"
                /db_xref="taxon:746"
                <1..153
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAB93470.1"
                /db_xref="GI:2731431"
                /translation="AFEFACSGGKIIRPGVPLGGTVVIAFADPDGKIEFTENKDA
                KAAALGN"
                237..887
                /codon_start=1
                /transl_table=11
                /product="RNase T"
                /protein_id="AAB93471.1"
                /db_xref="GI:2731432"
                /translation="MGOTTEQIDYKLLKHFRRGYPVIVDVFAGNNAQDALLEAA
                ITLKMPDQGYLVDPQCHFIAPFEGANINPESLKFNGIDIDNPLRGVAENAIIFGM
                FIVRKAMEQCGQGVIVAHNAASDPGFYRAAKIRAKRPFPPFAMFDFATLADG
                FMYGQTVLVKACQVAKIPFDGKQAHSAlyDTERTAELFCAMVNRKLDLGGFPPLSD"

```

CDS

```

1309..3063
/codon_start=1
/transl_table=11
/product="transferrin binding protein B"
/protein_id="AAB93472.1"
/db_xref="GI:2731433"
/translation="MFKLASSVLLNAALLAACSSNGSGSFDVQSAKVESQOTOTPKKP
SQDDNSNARRVVSASETALLQDGRGFSAKIPRRNLLPGQKDEVPKPEIKETDGL
PKPIEEEVKACGSADGESHALPHRKHLYTRDFNRSGYVYVHSGAPKPEIKELIR
TGAHGVYVYLGIEPKAFTQKTYGKYWDFTYAAKGRDSNIFLIPAGINSAGPEN
SHDINVDSEKPMGTGETADFANKLTGTVLRNGYVSRSEKQITTYIDIAKIG
NRFSGANPKPMILIFGKSTLEGFFGGAQLAGFLADKGVFVYFAGTRDAK
KDDSESADAFPIKLDKLSMDTFGNATHLIINNKQIPLTAETAKSFAEMFDDLV
TRIDGKTYRVSCCNLDYVFGIYSEGNNSDTALQELVGCERTALADLPTGVYR
GWGVMYKSGSAGVESGSRSLDFVFNKRLINGKLIANDGVVEERPMFLIE
GNLKGNGFGGTAKTNGSNGFNLDPKSTNGTGVGHINTQFEGGFGPKATELGGIVQTE
TKDRVSIITFGGKRQIEK"
3151..5943
/codon_start=1
/transl_table=11
/product="transferrin binding protein A"
/protein_id="AAB93473.1"
/db_xref="GI:2731434"
/translation="MIMKYHFRYSPVLTVLFPALSHSYGAATENKKIENNDAVLID
EVIVTSHVAHERONEVTGLGVKNYHMSKNQIIGIRDTRYDPGISVVEQGRGAS
SGYARGVDKNRVSLVLDLPQAHSHYTLGSDANGGAINIEIYEYENRSTELSKGASA
EYSGAHGGAIGFRTKDAODIIRKEQHWGLDSDKTSYASKNSHFLQSLAAGEAGFEA
LVIAHRHGKETKIHEANKLRIRITGFENRYDTQIPHRMLLEDLLIIVEDTCP
LDCTPRARVKLRNPNFPVPTPEYTPPEERKQLEQPYRTEQLSAQEYTGKRIAPNP
LDYKTSVFMFGYHFNSSHLYGAILEDTKYDIDRMOTPAYTKDDNLNLSRVVY
EGDNLLDGLVFXPRIPYGLRYSHVFFEDRHHKRLGFTYKVKPENRWLDSIKLSAD
KODIELYRLHLHCSDPYVDVKNCRPTLDKSWMYTERNYQEHRLVHLFEDKAL
NAGQGVFNQTHLKLGDYDFRNSLMDHGMDTAQYTKGTSYRGRDLNPIYRRD
PRSIETVSLCNRTRGDLNCEPRKIGKDSHFVSRDLVISEYVDLGLGRVDQGRFS
DDPMTLSRTYRNWNGGTLKPTPEVSLSYRISNGFRVPAFYELGKRDHIGLKONE
YVQRAORSHQLEPEKSTNHEIGVSPKQGYLDVSYFRNNYKMMIATACKRIIOKSHC
YVYHNIQDVALNGINLAKFDLHGILMLPDGFSYSAVNRVYKVKERKLTDSRLDSV
NPDLDAIQARYVLGFGDHPDEEKWGIGITTYTSKAKNADEVAGTRHHGHIHRVDLGG
KTGWSYTHDITGYINIKYNTLUGGIYVNTNRKYSTWESVRSQSVNAVNDGRGSNTRY
FCAGRNFSLAEMKF"
complement(6044..6340)
/notes="similar to factor for inversion stimulation
protein"
/codon_start=1
/transl_table=11
/product="FIS"
/protein_id="AAB93474.1"
/db_xref="GI:2731435"
/translation="MLEQQPTONPLVTMLNAQAOVKNPLRDNVKAALKNYLSQLNG
EPTLEYELVLSIEIHPMLDMYQYTRGQTRATMLGINRGTLRKLKKGNG"

```

CDS

```

BASE COUNT 1977 a 1132 c 1382 g 1883 t 4 others
ORIGIN

```

CDS

```

Query Match      99.6%; Score 2781; DB 2; Length 6378;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2792; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 atgataagaataatcatcttttcgcatttcacacctgttgcttaacagtgattattgct 60
|||||
Db 3151 ATGATAATGAATATCATCATTTTCGCTATTCACCTGTTCCTTTACACAGTGTATTGCT 3210
|||||

Qy 61 cttctcattcatcgcggtgctgcactgaaataataaaatcgaagaaataacgatcta 120
|||||
Db 3211 CTTTCTCATTACACGCTGCTGCGACTGAAATAAAAAAATCGAAGAAAAATAACGACTA 3270
|||||

Qy 121 qctgttcgtgaagattatgtgacagagaccattatgctcacgaacgaaacgaa 180
|||||
Db 3271 GCTGTTCTGGATGAAGTATTGTGACAGAGAGCCATTATGCTCACGACGTCACAAACGAA 3330
|||||

Qy 181 gtaactgcttgggggaagtagtgaaaaattatcacgaaatgagtaaaatcgaattctt 240
|||||
Db 3331 GTAACGGCTGGGGAAGTAGTGAAAAATATCACGAAATGAGTAAAAATCAAAATCTCT 3390
|||||

```

Qy 241 ggtattcgtgatttaactcgtctatgacctgtgtatttcggtggtggaacaagggtcgcggt 300
Db |||||
Qy 301 gcaagttagtgcctatgccattcgaaggttagataaaacccgtgccaagttacttgat 360
Db |||||
Qy 3451 gcaagttagtgcctatgccattcgaaggttagataaaacccgtgccaagttacttgat 3510
Db |||||
Qy 361 gggctaccacaagcgacagtttatcatcagctaggttcagatgcttaagtggtggtgcaatt 420
Db |||||
Qy 3511 gggctaccacaagcgacagtttatcatcagctaggttcagatgcttaagtggtggtgcaatt 3570
Db |||||
Qy 421 aatgagattgagatgaaacattcgttcaattgagttgaagaaaggagcaagttctcgc 480
Db |||||
Qy 3571 aatgagattgagatgaaacattcgttcaattgagttgaagaaaggagcaagttctcgc 3630
Db |||||
Qy 481 gaatatgctctggtgcgcatggtggtgctattggtttctgactaaagatgcgcagat 540
Db |||||
Qy 3631 gaatatgctctggtgcgcatggtggtgctattggtttctgactaaagatgcgcagat 3690
Db |||||
Qy 541 attattaaagagggcgacgattgggcttagatagtaagacaccttattgcccagcaaaaat 600
Db |||||
Qy 3691 attattaaagagggcgacgattgggcttagatagtaagacaccttattgcccagcaaaaat 3750
Db |||||
Qy 601 agccattttttacagttctatcgcagcggtggtgagggggtggttttgaagcacttggt 660
Db |||||
Qy 3751 agccattttttacagttctatcgcagcggtggtgagggggtggttttgaagcacttggt 3810
Db |||||
Qy 661 attgcaactcaccgacagcgttaagagaccacaaattctatccgagcaaaaataaataaa 720
Db |||||
Qy 3811 attgcaactcaccgacagcgttaagagaccacaaattctatccgagcaaaaataaataaa 3870
Db |||||
Qy 721 cataatattcggcgtataaccggcttgaaatcgctacgactttaccacaaattccgcac 780
Db |||||
Qy 3871 cataatattcggcgtataaccggcttgaaatcgctacgactttaccacaaattccgcac 3930
Db |||||
Qy 781 agaagtcctctggaggtatcctctttaaattggtgaagatacttggccacattagattgt 840
Db |||||
Qy 3931 agaagtcctctggaggtatcctctttaaattggtgaagatacttggccacattagattgt 3990
Db |||||
Qy 841 actcctcgtcaaggggttaagtgaacgcgcgataaattcccagtgagaaacatttcccgaa 900
Db |||||
Qy 3991 actcctcgtcaaggggttaagtgaacgcgcgataaattcccagtgagaaacatttcccgaa 4050
Db |||||
Qy 901 tatcgcctgaagcgcacacagcttgacagattccttatcgcactgacgactctca 960
Db |||||
Qy 4051 tatcgcctgaagcgcacacagcttgacagattccttatcgcactgacgactctca 4110
Db |||||
Qy 961 gcccaagaatataccggttaaagatcgcttgcaccaaaccctttagattacaagagtaat 1020
Db |||||
Qy 4111 gcccaagaatataccggttaaagatcgcttgcaccaaaccctttagattacaagagtaat 4170
Db |||||
Qy 1021 tctgttttatgaagtttggctatcaactcaactcgtctcattatcttggcgcaatctta 1080
Db |||||
Qy 4171 tctgttttatgaagtttggctatcaactcaactcgtctcattatcttggcgcaatctta 4230
Db |||||
Qy 1081 gaagatacaaaaacgcctacgatatcgctatgctatgcaaacccagcttactatatacaaaa 1140
Db |||||
Qy 4231 gaagatacaaaaacgcctacgatatcgctatgctatgcaaacccagcttactatatacaaaa 4290
Db |||||
Qy 1141 gacgatattaacttactcacttaggaactatggtttatgaaggggataaataatttttagatggc 1200
Db |||||
Qy 4291 gacgatattaacttactcacttaggaactatggtttatgaaggggataaataatttttagatggc 4350
Db |||||
Qy 1201 ttagtgttcaagccaagatcccttatggttggctatagccatgtaagttttttgat 1260
Db |||||
Qy 4351 ttagtgttcaagccaagatcccttatggttggctatagccatgtaagttttttgat 4410
Db |||||
Qy 1261 gaagtcacccaacacgcgtttaggattccctataataataaaccagagaataatcgc 1320
Db |||||
Qy 4411 gaagtcacccaacacgcgtttaggattccctataataataaaccagagaataatcgc 4470
Db |||||

Qy 1321 tgggttgatagcattaaactcagtcgcatgataaaacagataattgaactatatagccggcta 1380
Db |||||
Qy 4471 tgggttgatagcattaaactcagtcgcatgataaaacagataattgaactatatagccggcta 4530
Db |||||
Qy 1381 catcgttgcatttagcagattatccttggttagataaaaaatttccgcccacatttgat 1440
Db |||||
Qy 4531 catcgttgcatttagcagattatccttggttagataaaaaatttccgcccacatttgat 4590
Db |||||
Qy 1441 aaattcgttgcatttagcagattatccttggttagataaaaaatttccgcccacatttgat 1500
Db |||||
Qy 4591 aaattcgttgcatttagcagattatccttggttagataaaaaatttccgcccacatttgat 4650
Db |||||
Qy 1501 catttagaatttggataaagcgcataaactgctggttcaagcgctatttaacccaaacccacaaa 1560
Db |||||
Qy 4651 catttagaatttggataaagcgcataaactgctggttcaagcgctatttaacccaaacccacaaa 4710
Db |||||
Qy 1561 ctgaatttaggggttgggcttggattcgatttaattcgcttattggatcatcgtgggatatgact 1620
Db |||||
Qy 4711 ctgaatttaggggttgggcttggattcgatttaattcgcttattggatcatcgtgggatatgact 4770
Db |||||
Qy 1621 gcccaatatataccaaagcggttataccagctaccgcggttagagcggttagataaatacca 1680
Db |||||
Qy 4771 gcccaatatataccaaagcggttataccagctaccgcggttagagcggttagataaatacca 4830
Db |||||
Qy 1681 tataattatccgcgcgctatccacgcagctattgaaacggttatcttggtaataatacacgc 1740
Db |||||
Qy 4831 tataattatccgcgcgctatccacgcagctattgaaacggttatcttggtaataatacacgc 4890
Db |||||
Qy 1741 ggcagactcttaaacctgtaaacgcgcgttaaaattaaagcgctagccatttggtagcttc 1800
Db |||||
Qy 4891 ggcagactctcttaaacctgtaaacgcgcgttaaaattaaagcgctagccatttggtagcttc 4950
Db |||||
Qy 1801 cgcgacttagtgaataagcagctatggttggattggttaggttaggttgcgttttgcatacaat 1860
Db |||||
Qy 4951 cgcgacttagtgaataagcagctatggttggattggttaggttaggttgcgttttgcatacaat 5010
Db |||||
Qy 1861 cgatttaaatctgatgcctgacacacttagccgaacttatcgaaattggtcttgggaat 1920
Db |||||
Qy 5011 cgatttaaatctgatgcctgacacacttagccgaacttatcgaaattggtcttgggaat 5070
Db |||||
Qy 1921 ggtgggattacgcttaaacccacagagtttggattcgcttcttcttgcgaatttcaaacggt 1980
Db |||||
Qy 5071 ggtgggattacgcttaaacccacagagtttggattcgcttcttcttgcgaatttcaaacggt 5130
Db |||||
Qy 1981 tttagagctcctgactctatgaacttatggttaaacgctgactatattggctttaagat 2040
Db |||||
Qy 5131 tttagagctcctgactctatgaacttatggttaaacgctgactatattggctttaagat 5190
Db |||||
Qy 2041 aacgaatatgtgcaacgcgcgcaacgtagccacagtttagagccagaaaaatcgactaat 2100
Db |||||
Qy 5191 aacgaatatgtgcaacgcgcgcaacgtagccacagtttagagccagaaaaatcgactaat 5250
Db |||||
Qy 2101 catgagattggagtagctttaaggttcaatttggttacacttggatgagctatttcogt 2160
Db |||||
Qy 5251 catgagattggagtagctttaaggttcaatttggttacacttggatgagctatttcogt 5310
Db |||||
Qy 2161 aataactataaaaatgattgcacagctatgaagaaataatacaaaaaatcacactgt 2220
Db |||||
Qy 5311 aataactataaaaatgattgcacagctatgaagaaataatacaaaaaatcacactgt 5370
Db |||||
Qy 2221 ttctataactaccataatattcaagatgtagcactaaacgggataaaatttagtcgctaaa 2280
Db |||||
Qy 5371 ttctataactaccataatattcaagatgtagcactaaacgggataaaatttagtcgctaaa 5430
Db |||||
Qy 2281 tttagactaacggtatttttatctatgctgccagatggtttttttatcatcagttgcttat 2340
Db |||||
Qy 5431 tttagactaacggtatttttatctatgctgccagatggtttttttatcatcagttgcttat 5490
Db |||||
Qy 2341 aaccggttaaaagtaaaagagcggaataacacactaacgactcgatagcgtaaaagat 2400
Db |||||
Qy 5491 aaccggttaaaagtaaaagagcggaataacacactaacgactcgatagcgtaaaagat 5550
Db |||||
Qy 2401 cctattctagatgcgattccagccagcgcgtatgtgcttggattcggctacgatacacca 2460
Db |||||

Db	5551	CCTATTCTAGATCGATTACGCCAGCAGCTATGTGCTTGGATTGCGCTACGATCACCCA	5610	
Qy	2461	gaagaaataatgggaattggaattactactaccacccattctataagccaaacacgcgatgag	2520	
Db	5611	GAAGAAAATGGGAATGGCATTAACCAACCTATTCTTAAGGCAAAAGCCGCGATGAG	5670	
Qy	2521	gtgcaggcacacgtcatcacgg-atacatcgcttgatttagtggcaactaccggt	2579	
Db	5671	GTGGCAGGCACATCATCAGCGNATACATCGCTTGATTTAGTGGCAAACTGACCGGT	5730	
Qy	2580	tcttggtacacccatgatattaccggtttacatcaattataaaactacacacttcactgga	2639	
Db	5731	TC TTGGTACACCCATGATATTACCGGTTACATCAATATATAAACTACACCTTACGTGA	5790	
Qy	2640	ggaattataatgactaactcgtataatattccacttgggaatcagtcgccaatccggt	2699	
Db	5791	GGAATTTATAATGTGACTAAATCGTAAATATATCCACTTGGGAATCAGTCCGCCAATCCGGT	5850	
Qy	2700	gtgaatcagtaaaccaagacgggtgtagcaattacactcgatttgcgctccgggaga	2759	
Db	5851	GTGAATGCGGTAAACCAAGACGGGTGAGCAATTACACTCGATTGGCGCTCCGGGGAGA	5910	
Qy	2760	aatttcagtttagcatttgaatgaagttttag	2792	
Db	5911	AATTTCAGTTAGCATTTGAAATGAAGTTTATAG	5943	
RESULT 2				
APU16019				
LOCUS				
DEFINITION				
Actinobacillus pleuropneumoniae transferrin binding protein 2 (tbpB) gene, partial cds and transferrin binding protein 1 (tbpA) gene, complete cds.				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
1. .3541				
/organism="Actinobacillus pleuropneumoniae"				
/strain="H171"				
/db_xref="taxon:715"				
/clone="pBAPT1"				
/note="serotype 1"				
3. .707				
/gene="tbpB"				
<3. .707				
/gene="tbpB"				
/codon_start=1				
/transl_table=11				
/product="transferrin binding protein 2"				
/protein_id="AAC43486.1"				
/db_xref="GI:1055213"				
/translation="IDATKIDLTQFNKELNFGDASVLIIDGQKIDLAGVNFKNST				
VEINGKTMVAACCSNLEYMKFGOLWQKEGQKQVKNLSFLQGERATIDKMPAGGNTK				
YVGTWDLVSKGTNWIADNRRSSGYRTDFDNFSDKRYNGKLFDRGGVNPFTVDA				
TINGNGFIGSAKTSDSGFALDAGSSQHGNAVFSDIKVNGGFGYPTAGELGGQFHKSD				

NGSVGAVFGAKRQIEK"				
721. .3516				
/gene="tbpA"				
721. .3516				
/gene="tbpA"				
/codon_start=1				
/transl_table=11				
/product="transferrin binding protein 1"				
/protein_id="AAC43487.1"				
/db_xref="GI:1055214"				
/translation="MKNKLNLSLALLSFVQSYAEQAVQLNDVYVGTGKKAHKKE				
NEVTGLGVKVPDPTLSKQVILGIDRLTRYDPSVVEQGRGATGYSIRGVRNRVQ				
LALDGLPQIQSYVQYSSRSSGAINIEVENLRSIOISKGASSSESGSLGSGVQFR				
TKEVSDIIPKQSGWGLDTRKAYSSKNQOMLSLAFAGTHNGFDALVYTHRDGKETKA				
HDAESRSQNIIRVGVETNELDTSNRYTATTNNQHYGWFELKDECPDLDCTPKQVAR				
WKDTPSFRSYPEITPEKQAYENQKHITERLNAQDYTGAYRALPPLKYSQSLWLVK				
LGYTFSPKHVAGTYEHSQRDYTRDMTYTAYQPSDLLTRGNWYPMNNAKGLYRDN				
ALDGVADYFTEDGVKSSKGLRWAKARFIDEWHTRLGALRYTTNQDGNRLDRLSL				
SFDQKINLSRLRENNCSYPTIDKNCRATLDKLSWSTKNEOSSVEEKHDTIQLSLD				
KTVOQGLGHOLNMLLGSDFNSTLKRHELLSEFVSQVTHRIRGNQYKDTPIYELKD				
QAIYSKNECDYSCTIAGRADCATSKIGHNHXIALRDNEAITYLDIGLYRDKHKE				
RSTHRWANQGDINKSAWNIQIVAKPTSFSLSLSTRASSGFRVPSQELFGURYDQAMKG				
SSDAYQKTEKLSPEKSLNQEVAAATFQDFGVVEVSFKNDYKOLIIAPAEKMHQTQSMI				
NYFNVDQIKLDGINLIGLDQWNGVFKIPEGIYITTLAYSKMRVKEVNTQGYNIRSP				
LLDTIQARYVVGVDQDQEKWGNLTMTSHSSGNPNELRGNEQVGFANYERTATKK				
RTRSWHTFDLTGYITPWKHTTVRAGVYLMNYRYTTWESVQSSLNALHOHTNVKDYA				
RYAAPGRNYVVSFEMKF"				
BASE COUNT 1173 a 626 c 759 g 983 t				
ORIGIN				
Query Match 19.0%; Score 531; DB 2; Length 3541;				
Best Local Similarity 53.5%; Pred. No. 2.4e-114;				
Matches 144; Conservative 0; Mismatches 1160; Indels 100; Gaps 12;				
Qy	158	atgctcacgaacgtcaaaacgaagtaactggctggggaaagtagtgtaaaaattatcacg	217	
Db	836	AAGCACATAAAAGAGACGAGTCACAGCTTAGGGAAGTAGTTAAAACACCATATA	895	
Qy	218	aaatgagtaaaaaatcaaatctcttggtattcgtaattaaactcgctatgaccttgattt	277	
Db	896	CTCTTAGTAGGAGCAAGTGTTAGGAATACGAGATCTGACTCGTTACGACCCCGTATT	955	
Qy	278	cggtggggaacaagtcgcggtgcaagtagtgctgctccattcgaggtgtagataaaa	337	
Db	956	CTGTCGTAGAACAAAGGAGAGGTCGACTACAGCTACTCAATTCGGCGGGGTAGTCTA	1015	
Qy	338	accgtgtcagcttacttggttggtggctaccacaagcgacagttatcatcacgtaggtt	397	
Db	1016	ATCGTGTGGCTTGGCATTAGACGGTTTGCACAGATTCATCTATGTAAGCCCAATATT	1075	
Qy	398	cagatgctaattggtggtgcaattaatgagattgagtagtataaaacattcgttcaattgagt	457	
Db	1076	CACGTTCTCTCAAGCGGTGCCATTAAATGAAATAGAAATACGAAAAATCTCGGTCGATCA	1135	
Qy	458	taagcaagagagaagtcctcggaataatgctctggtgcgcatggtggtgctattggtt	517	
Db	1136	TTAGTAAGGGGCTGAGTCTCTCTGAGTTTGGTAGTGCTCCTACTAGGCGGTTTCGGTGC	1195	
Qy	518	tctgtactaaagatgcgcaggatattattaaagaggggcgagcatttggggcttagatagta	577	
Db	1196	TCCGTACCAAAAGAGTAAGCGACATTATTAAAGCCAGGCAATCTTTGGGATTAGATACCA	1255	
Qy	578	agacctcttatgcagcaaaaaatagccattttttacagtcctatcgacgcgctggtgaggg	637	
Db	1256	AAAGTGCTACAGTAGCAAAAATCAACAATGGTTAAACTCACTTGTCTTTTCGGGGTACTC	1315	
Qy	638	cggtgtgtttgaagcacttgatttgcactcaccgacacgcggttaagagagaccataatc	697	
Db	1316	ACAATGGCTTTGATGCTCTCTGTGATTTACACTACCGTGATGGTAAGAAACGAAAGCTC	1375	
Qy	698	atcccgaggcaataataataacataaatattcgccgctat-----aacccgcttgg	748	

Db 1376 ATAAAGATGCAGAGAGCGTTCTCAGAAATATCACCCGAGTAGGAGTGGAACCAACAGCAGC 1435
Qy 749 aaatcgtcagcaacttaccocaaattccgcacagaatgctcctcgtgagagatctctttaa 808
Db 1436 TTGATACCTCAATAGATATATCTGCGACGACGGAATAATCAACATACATTATGGCTGGTTTT 1495
Qy 809 ttgfggaagataacttgcccaacattagattgtaactcctcgtgcaagggttaagtgaacc 868
Db 1496 TGATTAAGATGAATGTCCAACGTTAGATTGTACGCCGAACACAGATGGCTAGGGTGACAA 1555
Qy 869 gcgataatttcccgagtgagaacatttccggaatatatacgcctgaagagcgcaaacagcttg 928
Db 1556 AAGATAGCCCATCTTCCGTTCTTACCTGAAATATATCTCTGAGGAAACAGAGCCTTATG 1615
Qy 929 agcagattccttctgcagctgagcagctctccagcccaagaatatatacccggttaagatcgca 988
Db 1616 AGAACCAAAAACATATTACAGAGCGTCTAAATGCTCAGGATTACACTGGTGAATATAGAG 1675
Qy 989 ttgcacaaacccctttagattacaagagtaattctgtttttttatgaagtgttgctatcact 1048
Db 1676 CTTTACCTGATCCCGCTTAAATATATAATCTGATTCTTGGCTGGTTAAATTTAGGATACACAT 1735
Qy 1049 tcaactgctcattatcttggcgcaactcttagaagatacaaaaaacacgcctacgatatcc 1108
Db 1736 TCTCTCCGAACATATATGCTCGCTGGTACTTTATGAACATAGCAACACGCGTTACGACACC 1795
Qy 1109 gtgatatgcaaacgccagcttac----- 1131
Db 1796 GAGATATGACCTATACCGCTTATTGGCAACCATCGGATTTACTTTAGAACCTGGTAGAAT 1855
Qy 1132 --tatacaaaagacgataattataactatacacttaggaactattgtttatgaaggggataata 1189
Db 1856 GGTATCCAATGAATATGCTAAAGGATATATCTGTAATAATGCTTTTAGATGGTGTGCTA 1915
Qy 1190 ttttagatggcttagtgttcaagcgaaggtcccttattgggttgcgtatagccatgtga 1249
Db 1916 TTGACTACTTTACCGAGATGGTGTGAATCATCAAAAGGTTTACGTTGGCCAAAAGCTC 1975
Qy 1250 agttttttgtaagcgtcaccacaaacgtgcttttagattacacatataaatat---aaac 1306
Db 1976 GTTTTATTGAGGAGTGGCACACTCGTGATCGCTTAGGTGCTTTATATCGTTTATACCAATC 2035
Qy 1307 cagagaataatcgtgttgatagcattaaactcaactcagtgcggtgataaacaagattgaac 1366
Db 2036 AAGATGGAATTCGTGATTGATAGACTATCTTTCAGTTTCGATCAGCAAAAAATTAAT 2095
Qy 1367 tatatagcggctacatcgtcttgatgtagcgtattactcgttggttagataaaaattgcc 1426
Db 2096 TATCTACCGCTTGAGAGAGAACAACTGTTCGGAATATCCAAACCATAGATGAAGATTGCC 2155
Qy 1427 gccgactttggataaaacttggctctatgtatcgaaactgagcgtgaataattaccagaata 1486
Db 2156 GTGCAACTCTTGATAAACTTTGGCTTCAACTAAAATGACCAAAAGTCTTATGAAGAA 2215
Qy 1487 agcatggtgctatcatttagaattttagataaagcgtataaactgctggtgtaagggcgtat 1546
Db 2216 AACACGACACTATTACGCTCTCGTTAGATAAAACCGTACAAACGGGATGGG----- 2267
Qy 1547 accaaaccccaactgaatttaggggttggttggcttggatgatttaattcgtttggtatc 1606
Db 2268 ----TAAACATCAATTAATATGTTATTAGGTTTACAGACCGTTTCAATTCACCTTAAAC 2323
Qy 1607 atggggatgactgcccataataccaaagcggttataccagctaccgcggttagagggc 1666
Db 2324 GCCAGAAATTTTGGTGAATTTCTGTGGGAACCTTGGCATCGTATCAGAGGTAAACGGTT 2383
Qy 1667 gtttagataatccatataatttatgcgc-----gcgatccacgcagattgtgaacgggtat 1720
Db 2384 ATAAAGATACACCTTACATCTATGAGCTATAAAGATCAGGCAATTTATAGTAAAAATGAAT 2443
Qy 1721 ctttgtgtaataatacacgcggcgacatctttaaactgtgaaccgcgttaaaattaaagcg 1780
Db 2444 GTGATTATAGTGGCACTATTTCAGGTAGGGCTGATTGTGCTTACAAGTAAAAATCAAGGGC 2503

Qy 1781 atagccattttgttagcttccgcg---atctagtataaagcgagtgatgtggtattgggat 1837
Db 2504 ATAATCACTACATCGCTCTGAGAGATAAATTTGCCATAACAAGTATTTTGGATATTGGTT 2563
Qy 1838 taggggtgcgttttgatcaacatcgattttaaactctgatgacccgtggacactgagcgaa 1897
Db 2564 TGGGTTACCGTTTCGATATAGCAATAAATTCGATGACACTCATCGTGGGCAAAACAAGCG 2623
Qy 1898 ctatcgaaattggcttggaaatgggtgggtattacgcgttaaaccacaacagagttgtatcgc 1957
Db 2624 ATTATAAAACACGTGCGTGAATATTGGCATAGTCGCAAAACCAACGTCATCTCTATCGC 2683
Qy 1958 ttcttctatgcatttcaaacggttttagagtgctcgtcattctatgaactttatggttaac 2017
Db 2684 TCTCTTATCGAGCATCATCTGGCTTTAGAGTGCCAAAGTTTCCAAAGAGCTATTGTGGCTTAC 2743
Qy 2018 gtgatcatattgggcttaaaagataaagaaatatgtgcaacgcgcgcaacgtagccaccagt 2077
Db 2744 GTTATGATGTGCAATGAAAGGCTCCAGCGATGCTTA-----CCAAAAACAGAGAAGT 2797
Qy 2078 tagagccagaaaaaatgcactaatcatgatgatgtgagttagctttaaagggtcaatttggtt 2137
Db 2798 TATCTCTGAAAAATCCTTAACCAAGAGGTTGCTGGCACTTTCAAAAGGTGATTTTGGTG 2857
Qy 2138 acctgtgtgagctatttcccgtaactataaaaaatatgattgcgcagcagcatgtataaa 2197
Db 2858 TCGTTGAAGTCAGTTATTTCAAAAAATGACTATAAGCAGTTAATTGCTCCAGCAGA----- 2912
Qy 2198 gaataatacaaaaaacacactgtttctataactaccataataattcaagatgtagcactaa 2257
Db 2913 -AAGATGCAATCAAACTCAATCAATGATTAACTATTTTAATGTCGAAGATATTAAATTTGG 2971
Qy 2258 acgggataaaatttagtcgctaaatttgaacttacacggtattttctatctatcgtgccagatg 2317
Db 2972 ACGCATATATCTTATTTGGTTAAGCTAGATTGGAATGGGTTATTGTGATAAATTTCTCAGG 3031
Qy 2318 gtttttattcatcagttgcttataaacggtgtaaaagtaaaagcggaagcggaactaacccgact 2377
Db 3032 GCATTTACACAACATTTGGCTTATAGCAAAATGCGAGTAAAGAGGTGA---AAACTATC 3088
Qy 2378 caagactcgtatagcgttaaacgactcctattctagatggattcagccagcagcgtatgac 2437
Db 3089 AAGGGTATATGAATATTTCGTTCTCCATTGTTAGATACCATTCAGCCCTGCTCGCTATTGG 3148
Qy 2438 ttggtatcgcgtacgataccaccagaaagaaatggggaattggcattactaccacactatt 2497
Db 3149 TAGGAGTGGGTACGATCAGCCAGATGAAAAATGGGCGGTGAATCTAACAAATGACACACT 3208
Qy 2498 ctaaagccaaaaacgcgcatgaggtggcgacacgctcatcgaggtacacggtatcgctgtga 2557
Db 3209 CCAGTGGAAAAAATCCAAATGAGTTAAGAGGTAATGAACAAGTCGGTTTTCGCAATTTATG 3268
Qy 2558 ttttagtgcgc-----aaactgacggttcttggtcacaccatgatattacoggttaca 2610
Db 3269 AGCGAACTGCCAGAGAAAGAAACACGCTTTTGGCATACCTTTGACTTAAACGGGATATA 3328
Qy 2611 tca---attataaaactacaccttaccgtggagaaatttataatgtgactaaactgtaaat 2667
Db 3329 TCACCCCTTGGAAACATPACAAACGGTACGAGTGGCTGATATATACCTGATGAATTTATCGTT 3388
Qy 2668 attcacttgggaatcagtcgcgaatcccggttgtagtcagtaaaacacagacgcgg---g 2724
Db 3389 ACACCACTTGGGAATCCGTCAGTCAATCTTCGCTTAAATGCAATTCATCAGCATACTAACG 3448
Qy 2725 gttagcaattacactcgtatttggcgctccgggggaaatttcaagtttagcatttgaatatga 2784
Db 3449 TAAAAGACTATGCAAGGTATGACGCCGCCGTTAGAAAATTTATGTTTTCATTCGAAATGA 3508
Qy 2785 agtttta 2791
Db 3509 AATTCTA 3515


```
/codon_start=1
/transl_table=11
/protein_id="CAA89810.1"
/db_xref="GI:853709"
/db_xref="SPTREMBL:Q44166"
/translating="MKNKLNLIJSLALLSFVQSVAEQAVQVNDVYVTVTKKHAHKE
NEVTGLGVKVPDLSLKEQVIGIRDLYDPGIVSEQGRGATGYSIRGVDRNRVG
LALDGLPQIQSVQSVSRSSGAINIEYENLRSIQISGASSSEFGSGLGGSVQFR
TKEVSDIIPKPGSWGIDTSAYSSKNOHLSIAPAGTHNGFDALVYVTHRDKETKA
HKDAERSRNIIRVGVETNELDTSNRYTATNNQHTYGFHLKDCSPILDCPDKOMAR
VNDPSPFSPEYTPPEKQAYENKHITERNAQDITGEYRALDPLFLKYSDSLWK
LGYTSPFTYDTEHSHKQYRDYMTYAYQPSDLRTGRNRYPMNNAKGLYRDN
ALDQVAIDYFEDGVKSKGLRWAKARFIDEWHTDRGLGALRYNQDNRLIDRLSL
SPDOQRNLSTLRNENCSEYPTIDKNCRAITDKLWSSTKNEQSSVEEKHDPIQLSLD
KTVQCLGKHOLNMLLGSDFRNSLTKRHEILSEFSVGTWHRIRGNQYKDPYIYELKO
QAIYSKNECDYSGTTAGRADCAITSIKIGHNHIALDNFAITKYLDIGLGYRFRDKHF
RSTHWANGQDKNSAWNIGIVAKPTSFUSLSYRASSGPRVPSFOELFLKRYDGAAMK
SSDAQTKLEKSPKLSNEVAATKGFVVEYFKNDRYKQLIAPAPERMHQTSMI
NYFNVDIKLDGINKLDWNGVDFKIPEGYITTLAYSKMRVKEVKNYQVMYNTATRP
LLDTIQPARYVGVGVDQDEKQVNLITWSSGNPNELRNEQVGFANYBERTAKK
RTLSWHTFDLTGYIYTPKHKHTVRAGVYNLMNYRYTWESVRSLSLNAHQHNVKDYA
RYAAPGRNVYVSEMKF"
BASE COUNT      928 a      519 c      598 g      788 t
ORIGIN

Query Match      19.0%; Score 529.4; DB 1; Length 2833;
Best Local Similarity 53.4%; Pred. No. 5.8e-114;
Matches 1446; Conservative 0; Mismatches 1161; Indels 100; Gaps 12;

QY 158 atgtctacgaagtcgaacgaagtaactgcttggtgggaagtagtgaaaaattatcaag 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126 AAGCACATAAAAGAACGACGACAGCTACAGGCTTAGGGAAGTAGTAAACACCCAGATA 195

QY 218 aaatgagtaaaatcaaatcttctgtattctgtatttaactgcctatgaaccttgattt 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 186 CTCTTAGAGGACCAAGTGTAGGAATACGAGATCTGACTCGTTACGACCCCGGTATT 245

QY 278 cgggtgggaacaagtcgctgggtgcaagtagtgctatgccattcgaggtgtagataaaa 337
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 246 CTGTCGTAGAACAAAGGAGAGGTGCGACTACAGGCTACTCAATTCGGGGGTAGATCGTA 305

QY 338 accgtgcagcttacttctgtatgggtaccacaagcgacaggtatcatcacgctaggtt 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 306 ATCGTGTGGCTTGGCATTAGACGGTTTGGCCACAGATTCAATCTATGTAAAGCCATATT 365

QY 398 cagatgcataatggtgtgcaataatgagattgagatgataaaacattcgttcaattgagt 457
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 366 CACGTTCTCAAGCGGTGCCATTAATGAATAGAAATACGAAATACTGCGTTCGATCCAAA 425

QY 458 taagcaagagagaagttctcggaataatgctctggtgcgcatggtgtgctattggtt 517
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 426 TTAGTAAAGGGGCTAGTCTTCTGAGTTTGGTAGTGCTCCTACAGCGGTTTCGGTGCAAT 485

QY 518 ttcgtactaaagtcgcagagataattataaagggcgacagcattggggtcttagatagta 577
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 486 TCCGTACCAAGAGGTAAAGGCACATTATTAAAGCAGGCGCAATCTTGGGGATTAGATACCA 545

QY 578 agacctttagcagcaaaaatagccattttttacagtcctatcgacgctggtggtgag 637
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 546 AAAGTGCCTACAGTAGCAAAAATCAACAATGGTTAAATCATTGCTTTTTCGGGTACTC 605

QY 638 cgggtggttttgaagccttctgtattgcaactcacgcacgcggttaagagaccaaattc 697
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 606 ACAATGCGCTTGTAGTCTTCTGTGATTTTACACTCACCTCACCCTGATGTGAAGGAAACGAAAGCTC 665

QY 698 attccgaggaataataataataatattcgccgtat-----aacggccttg 748
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 666 ATAAAGATGACAGAGTCGTTCTCAGAAATATCACCCGAGTAGAGMGTCGGAACCAACAGC 725

QY 749 aaatcgtctacgactttaccacaaatccgcacagaatgctcctcgtgagatctctttaa 808
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 726 TTGATACCTCAAAATAGATATACTCGACAGCAGGAATAATCAACATACCTATTATGGCTGGTTT 785
```

```
QY 809 tigtggaatacttgcacacattagattgtactctctgctcaaggggttaagttaacc 868
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 786 TGATTAAGATGAATGTCCCAAGTTAGATTGACGCCGAAACAGATGCGTAGGTGACAA 845

QY 869 gggataatttcccagtgagaacatttcccgggaatatatacgcctgaagagcgcaacagcttg 928
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 846 AAGATACGCCACTCTTTCCGCTTCTTACCCTGAATATATACTCTGAGGAAAAACAGGCTTATG 905

QY 929 agcagattccttatcgactcagtcagtcagtcctcagcccaagaataataccggttaagaatcgca 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 906 AGAACCAAAAAACATATTACAGAGCGCTCTAAATGCTCAGGATTACACTGGTGAATATAGAG 965

QY 989 ttgcacaaaccccttagatttcaagagaaatctgttttttgaagtttggctatcaact 1048
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 966 CTTTACTCTGATCCGCTTAATATAATCAATCTGATTCCTTGGCTGGTTAAATTAGGATACACAT 1025

QY 1049 tcaactcgtctctattatcttggcgcaatcttgaagaatacaaaaaacacgcctacgatccc 1108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1026 TCTCTCCGAAAAACATTATTGTCGCTGGTACTTATGAACATAGCAAAACAGCGTTACGACACCC 1085

QY 1109 gtgatgtcaaacgccagcttac----- 1131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1086 GAGATATGACCTATACCGCTTATTGGCAACCATCGGATTCTTAGAAGTGTAGAAATT 1145

QY 1132 --tatacaaaagacgatattaaacttatcacttaggaactatgtttatgaagggataata 1189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1146 GGTATCCAATGAATAGCTAAAGGATTATATCGTGATTAATGCTTTAGATGGTGTGCTA 1205

QY 1190 ttttagatggcttagtgtttcaagcgaagatcccttatgggtgcgctatagccatgtga 1249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1206 TTGACTACTTTACGGAAGATGGTGTGAAATCATCAAAAGGTTTACGTTGGGCAAAAGCTC 1265

QY 1250 agttttttagtaaacctcaccacaaacgcgtgtttaggattcaacctataatat---aaac 1306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1266 GTTTTATTACGAGTAGGCACACTCGTGATCGCTTAGTGCTTTATATCGTTATACCAATC 1325

QY 1307 cagagaataatcgctggttgtagatgattaaactcagtcggtggaatacaagaatattgaac 1366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1326 AAGATGGAATCTCTGATGTATGATAGACTATCTTGTAGTTTCGATCAGCAAAAAATTAATT 1385

QY 1367 tatatagcgggtcacatcgcttgcatgtagcattatcctggtgtagataaaaaattgcc 1426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1386 TATCTACCGCTTTGAGAGAGAACCAACTGTTCCGAATATCCAACCATAGATAAAGATTGCC 1445

QY 1427 gccgactttgtaataaattggtctatgtatcgaaactgagcgtaataataccagaataa 1486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1446 GTGCAAACTCTTGATAAACTTTGGTCTTCAACTAAAAATGAGCAAAAGTTCTCTATGAAGAAA 1505

QY 1487 agcatcgtctcattcattagaatttgataaagcgtctaaatgctgctcaagcggtattta 1546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1506 AACACGACACTATTTCAGCTCTCGTTTAGATAAAACCGGTACAAACGGGATTGGG----- 1557

QY 1547 accaaacccacaaactgaatttagggttgggtctttagcatttaattcgtcttatggtac 1606
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1558 ----TAAACATCAATTAAATATGTTATTAGTTTCAGACCGTTTCAATTCACCTTAAAC 1613

QY 1607 atgggggatgactgccccataaccacaaagcggttatccagcactaccgcggttagagggc 1666
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1614 GCCACGAAATTTTGTGTAATTTTCTGTGGAACTTGGCATCTGATCAGAGGTAACGGTT 1673

QY 1667 gtttagataatccatatatttatcgcc-----gcgataccgcagctattgaacggtat 1720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1674 ATAAAGATACACCTTACATCTATGAGCTAAAAGATCAGGCAATTTTATAGTAAAAATGAAT 1733

QY 1721 ctttgtgtaataatacacgcggcgacatctttaaactgtgaacccgcgtataaattaaagcg 1780
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1734 GTGATTATAGTGGCATTATTGTCAGGTAGGGCTGATTGTGCTACAAAGTAAATCAAGGGC 1793

QY 1781 atagccattttgttagcttcgcg----atctagtgataagcgagtgatgtggatttgggat 1837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1794 ATAATCACTACATACGCTCTGAGAGATAAATTTTGGCCATAACCAAGTATTGTGATATTGGTT 1853
```

```
Qy 1838 taggggtcggttttgatacaacatcgatttaataatctgatgacggtgggacacttaagccgaa 1897
Db 1854 TGGGTTACGGTTTCGATAGCAATAATCCGTPAGCACATCGCTGGCGCAATCAAGCG 1913
Qy 1898 ctatcgaattggtcttggaatggtggattacgcttaaaccaacagagtttgatcgc 1957
Db 1914 ATTATAAACAAGTCGGTGGATATATGGCATAGTCGCAAAACCAACGTCATCTCTATCGC 1973
Qy 1958 ttcttatgcatttcaaacggttttagatgctgcctgattcttatgaacttttatgtaaac 2017
Db 1974 TCTCTATCGAGCATCATCTGCTTTAGAGTGCCAAAGTTTCCAGAGCTATTTGGCTTAC 2033
Qy 2018 gtagcatattgggtttaagaataacgaatatgtcaacgcgcgcaacgtagccaccagt 2077
Db 2034 GTATATGCTGTCATGAAGAGCTCCAGCATGCTTA-----CCAAAAACAGAGAGT 2087
Qy 2078 tagagccagaaaaatcgactaatcatgatgagttgagttagcttttaaaagttcaattgatt 2137
Db 2088 TATCTCTGAAAAATCCTTAACCAAGAGGTGCTGCGACTTTCAAAGTGATTTTGGTG 2147
Qy 2138 accttgatgtgagctatttccgtaataactataaaaaatatgattgcgacagcatgtaaaa 2197
Db 2148 TCGTTGAAGTCAGTTATTTCAAAAATGACTATAAGCAGTTAATTGCTCCAGCAGA---- 2202
Qy 2198 gaataatacaaaaaatcaactgtttcttataactacataataattcaagatgtagcactaa 2257
Db 2203 -AAGATGTCATCAAACTCAATCAATGATTAACATATTTAATGTGCAAGATATAAATTGG 2261
Qy 2258 acggataaatttagctcgataattgacttacacggatttttattctatctgctgccagatg 2317
Db 2262 ACGGCATTAACTTATGTAAGCTAGATTTGGAATGGGGTATTTGATAAAATTCCTGAGG 2321
Qy 2318 gttttattcatcagttgcttataaccggtgtaaaagataaagagcggaactaacccgact 2377
Db 2322 GCATTTACACACATGCTTATACCAATGCGAGTAAGAGGTGAA---AACTATC 2378
Qy 2378 caagactcgatagcgtaaacgatcctatctatagatgcgattgaagccagcagctatgyc 2437
Db 2379 AAGGGTATATGAATATGCTTCCATCTGTAGATACCATTCAGCTGCTGCTATGTTG 2438
Qy 2438 ttggtatcgctacgatcacccagagaagaaatgggaattggcaattactaccacatt 2497
Db 2439 TAGGAGTGGGTACGATCAGCAGATGAAAAATGGGGGTGAATCTAACAATGACACACT 2498
Qy 2498 ctaaagccaaacacccgatggtggtgagccagcacgcgtacacggtacacatcggtga 2557
Db 2499 CCACTGGAAAAATCAATGAGTTAAGAGTTAATGACAGTCGCTTTGCCAATTATG 2558
Qy 2558 tttaggtggc-----aaactgacgggttcttgggtacacccatgatattaccgggtaca 2610
Db 2559 AGCGAACTGCCACGAGAAAGAACACTTCTTGGCATACCTTTGACTTAACGGGATATA 2618
Qy 2611 tca----attataaaactacacttacgtgaggaattataatgtgactaatgttaaat 2667
Db 2619 TCACCCCTTGGAAACATACACGGTACGAGTGGCGGTATATAACCTGATGAATTATCGTT 2678
Qy 2668 attccacttgggaatcagtgcccaatcggtgtgtaagtcagtaaaccaacacgg---g 2724
Db 2679 ACACCACTGGGAATCCGTACGCTCAATCTTCGCTTAATGCAATTCATCAGACTACTACG 2738
Qy 2725 gtagcaattacactgattggcgtccggggagaaatttcagtttagcatttgaatatga 2784
Db 2739 TAAAGACTATCAAGGTATGACGCGCCGGTAGAAATATATGTTTGTTCATTTCGAATGA 2798
Qy 2785 agtttta 2791
Db 2799 AATTCTA 2805

RESULT 5
APDNATFBB APDNATFBB 3447 bp DNA BCT 20-SEP-1995
LOCUS A.pleuropneumoniae tfbb gene encoding transferrin receptor.
DEFINITION
```

```
ACCESSION 254191
VERSION 254191.1
KEYWORDS GI:994790
SOURCE tfbb gene; transferrin receptor.
ORGANISM Actinobacillus pleuropneumoniae.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Actinobacillus.
REFERENCE 1 (bases 180 to 2873)
AUTHORS Wilke,M.
JOURNAL Thesis (1995) Institut fuer Mikrobiologie, Tieraerztliche
Hochschule Hannover
REFERENCE 2 (bases 1 to 3447)
AUTHORS Wilke,M.
JOURNAL Direct Submission
TITLE Submitted (19-SEP-1995) Wilke M., Tieraerztliche Hochschule
Hannover, Institut fuer Mikrobiologie, Bischofsholer Damm 15,
Hannover, Germany, 30173
FEATURES
source
1. 3447
/organism="Actinobacillus pleuropneumoniae"
/plasmid="pTF205/023"
/strain="AP205"
/db_xref="taxon:715"
/clone_lib="lambda 2001-library"
/notes="serotype:7"
180..2876
/gene="tfbb"
180..2876
/gene="tfbb"
/function="binding of porcine transferrin"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="transferrin Receptor"
/protein_id="CAA90896.1"
/db_xref="GI:994791"
/translaton="MKNLNLISLALISLFAVSQAQVQLNDVYVTGTTKKAHKE
NEVTGLKVVKTPDLSKSEQLVLRDITRYDPSIVVQGRGATTVSIRGVDNRVVG
LALDGLPOIOSVYVSRSSGAINETEVENLRSIOISKASSEFGSLGSGVQPR
TKEVSDIILKFGQWGLDTKSAYSSKNOUWNSLAFTHNGFESLVYITTHDKRETKA
HKDAESKSIQRVLDADNPQSNFVKVNDPCSLNCFQPGVQVSYDQKITEQL
SAKEYTEERALPDVVKYKSDLSWLKGLYSKPKHYGVYHSHKQVDRKMTPEY
WQSLDKNGDKWPMNNAKGLYRONALDGVADYFTEDGVKSKGLRWAKRFIDENH
TRORIALYRYTNODGNRLIDRLSLSPDOOKINLSTRLENNCSEYPTIDNCRATLD
KLWSPKNEQSSVEERHDTIQLSDTKVQGLCKHQLMLLSDRFNSTLKRIHELSE
FVSTWHIRNGIKDTPYIELDKQAIYSKNECDYSGTIGRADCATSKIKGNHII
ALRDNFAITKIDIGLYRFDKHKFRSHWANOGDYKNSAWNIGIVAKPTSLSLY
RASSGFRVPSFQLFLGRYDGMKMGSDAYQKTEKLSPEKSLNQEAATFKGDFGVY
VSYFKNDYKOLIAPAERHMQTSMINYNFVQDIKLDGINLGLDNGVDFDKIPEGY
TTLAYSKMRVKEVKNYQGYNNIRSPLLDTIOPARYVGVGVYDOPDEKQWNLMTTHSS
GKNPNELRGNEQVGFANYERTATPKTRSHHTFDLGYITPWKHTTVRAGYNNLMYR
YTTWESVRQSSLAIRHOHTNVKOTRAAPGRNVVVSFEKPF"
BASE COUNT 1160 a 610 c 719 g 958 t
ORIGIN
Query Match 16.2%; Score 452.2; DB 1; Length 3447;
Best Local Similarity 52.4%; Pred. No. 7.8e-96;
Matches 1393; Conservative 0; Mismatches 1163; Indels 103; Gaps 14;
Qy 158 atgtctcacgaacgtcaaaaacgaagtaactggtctgggaagtagtgaaattatcacg 217
Db 295 AAGCACATAAAAGAGACGACGAGCTTAGGGAAGTAGTGAACACACAGATT 354
Qy 218 aaatgagtaaaatcaaatcttctggtattctgatttaactcgctatgacccctgattt 277
Db 355 CTCTTAGTAAGGAGCAAGTGTAGGAATCGAGATCTGACTCGCTACGATCCGGTATT 414
Qy 278 cgggtgggaacaaggtcgcggtgcaagtagtgctatgccattcgagaggtgaataaaa 337
Db 415 CTGTAGTAGCAAGGAGGAGGTGCAACGACAGGCTACTCAATTCTGCTGGGTAGATCGTA 474
```


Db 2557 CAATGACACACTCCAGTGGAAATAATCAAAATGAGTTAAGAGGTAAATGAACAAGTCGGTT 2616
Qy 2546 acatgcgcttgattagtgagc-----aaactgaccggtctcttggtacaccatgata 2598
Db 2617 TTGCCAATTATGAGCGAAGTCCACGAGAAAGAACACAGTCTTGGCATACCTTGACT 2676
Qy 2599 ttaccggttaacata---attataaaactacacaccttactgctggaggaatttataatgtga 2655
Db 2677 TAACGGGATATATCACCCCTTGGAAACATACAAACGTCAGGCTGGCGTATATAACCTGA 2736
Qy 2656 ctaactcgaataattccacttggaatcagtcgccaatccggtgagtaacagtaaac 2715
Db 2737 TGAATATTCGTTACACCACTTGGGAATCCGTACGTCATCTTCGCTTAAATGCAATTCAT 2796
Qy 2716 aagaccgg---ggtagcaattacacactgattggcgctccggggagagaatttcagtttag 2772
Db 2797 AGCATACTAACGTAAGACACTATGCAAGGTATGCAGCGCCCGGTAGATAATATGTTGTTT 2856
Qy 2773 catttgaataagagtttta 2791
Db 2857 CATTCGAATGAATTCCTA 2875
RESULT 6
APU16017 5083 bp DNA BCT 09-NOV-1995
LOCUS Actinobacillus pleuropneumoniae transferrin binding protein 2
DEFINITION (tbpB) and transferrin binding protein 1 (tbpA) genes, complete
cds.
ACCESSION U16017
VERSION U16017.1 GI:1055209
KEYWORDS Actinobacillus pleuropneumoniae.
SOURCE Actinobacillus pleuropneumoniae
ORGANISM Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
Actinobacillus.
REFERENCE 1 (bases 1 to 4575; 1 to 1855)
Gonzalez,G.C., Yu,R.H., Rostek,P.R. Jr. and Schryvers,A.B.
TITLE Sequence, genetic analysis, and expression of Actinobacillus
pleuropneumoniae transferrin receptor genes
JOURNAL Microbiology 141 (Pt 10), 2405-2416 (1995)
MEDLINE 96036198
REFERENCE 2 (bases 1 to 5083)
Schryvers,A.B.
AUTHORS Direct Submission
TITLE Submitted (17-OCT-1994) Anthony B. Schryvers, Microbiology and
JOURNAL Infectious Diseases, University of Calgary, 3330 Hospital Drive NW,
Calgary, Alberta, T2N 4N1, Canada
FEATURES
source Location/Qualifiers
1. 5083
/organism="Actinobacillus pleuropneumoniae"
/strain="H49"
/db_xref="taxon:715"
/clone="pBAP11"
/note="serotype 7"
197. .200
206. .1849
/gene="tbpB"
206. .1849
/gene="tbpB"
/codon_start=1
/transl_table=11
/product="transferrin binding protein 2"
/protein_id="AAC43484.1"
/db_xref="GI:1055210"

GKNLSGKLFDRKNVNFVTPDKIDNGFTGAKTSDAGFALDSSGSSRYENVKFNDV
AVSGGFYGPPTAAELGGQFHHKSENGSVGAVFGAQKVK"
1852. .1855
1863. .4559
/gene="tbpA"
1863. .4559
/gene="tbpA"
/codon_start=1
/transl_table=11
/product="transferrin binding protein 1"
/protein_id="AAC43485.1"
/db_xref="GI:1055211"
/translation="MNKNLNLISLALSLFAVSQAQVQLNDVYVGTGKKAHKKE
NEYTGKVKVTPDSLSKEQVLGMRLTRYDPGISVVEQGRGATTGYSIRGVDRNVYG
LALDGLPOIQSVYVSOYSSSSGAINETIENLRSIOISKSGASSSEFGSGSLGGSVQFR
TKVSDIILKPGOSWGLDTKAYSCKNQWLSLAFAGTHNGFDALVLYTHRDGKETKA
HKDAERSKSIQRVLDLADNPNQSNWFKVNDPCSLNCEPQQVQSVYSDQKLTIEQL
SAKEYTEERSLPDPVKYKSDSLWLVKGLYSLSPKHYVAGVYEHSGKQYTRDMTTPAY
WQSLDLKNGDKWTVMNNAKGLYRDNALDGVADYFTEDGVKSSKGLRWAKARFIDEWH
TRDLGALYRYTNDQGNRLDLRLSLDFDQOKINLSTRLENNNGSEYPTIDNCRATLD
KLWSSTKNEOSVEEKHDTIQLSLDKTVOTGLKHOLNMLLGLSDRNFSTLKRHEILSE
FSVGTWHRIRGNGYKDTPIYELKDQALVSKNEDYSGTTAGRADCATSKIKGNHVI
ALRDNFAIKYLDIGLYRFDKHKFRSTRWANQGDYKNNAWNGIYAKTSTLSLSI
RASGFRVPSFQELFLYDGMKSSDAYQTEKLSPEKSLNQEAATFKGDFGVVE
VSYFKNYKOLIAPAERMHQTQSMINYNFVQDKLDGINLGLDWNKGVDFDKIPGEGY
TTLAYSKMRVKEVQNTQGYNNIRSPILDTIQPARYVGVGYDQDEKWKGVNLTWTHSS
GKNPNELRGNEOVGFANERTATKRTSRWHTFDLTCYITPMKHTITVRAGVYNLMNVR
YTTWESVRSQSSLAHIOHTNVNDYARAYAAPGRNVVVSFEMKF"
BASE COUNT 1727 a 857 c 1074 g 1425 t
ORIGIN
Query Match 16.1%; Score 450.6; DB 2; Length 5083;
Best Local Similarity 52.4%; Pred. No. 1.9e-95;
Matches 1392; Conservative 0; Mismatches 1164; Indels 103; Gaps 14;
Qy 158 atgtctcgaacgtcaaaacgaagtaactggtctgggaaagtagtgaataatcaccg 217
Db 1978 AAGCACATAAAAGAAAGAACGAGTACAGGCTTAGGGAAGTAGTAGAAACACCAGATT 2037
Qy 218 aaatgagtaaaaaatcaaatctcttggtattcgatttaactcgtatgaccttgattt 277
Db 2038 CTCCTTAGAGCAAGTGTAGGAATGCGAGATCTGACTCGCTACGATCCAGGTATT 2097
Qy 278 cgggtgggacaaggtcgcgctcaagtagtgccttcgattccgattcgaggttaataaaa 337
Db 2098 CTGTAGTAGCAAGAGAGAGAGTGCACGACGAGCTACTCAATCTCGTGGGTAGTCGTA 2157
Qy 338 accgtgtcagcttactgtgtggtctaccacaagcgacagttatcatcagctaggtt 397
Db 2158 ATCGTGTGGGCTTGGCATTAGACGGTTTGCACAGATTCATCTATAGTCAATATT 2217
Qy 398 cagatgctaatggtggtgcaattaatgagattgagttatgaataacattcattgagtt 457
Db 2218 CACGTTCTCTAAGCGGTGCCATTAAATGAATAGAAATACGAAAAATCTCGGTCGATCCAAA 2277
Qy 458 taagcaaaaggaagacttcgcggaataatggtctggtgcgcatggtggtgctattggtt 517
Db 2278 TTAGTAAGGAGCTAGTCTTCTGAGTTTGGCAGCGGCTCGCTAGGCGGCTTCGGTGCAAT 2337
Qy 518 ttctactaaagatgctcagcaggtattattaaagggggcagcattggggcttagatagta 577
Db 2338 TCCGTACCAAGAGGTAAAGCAGCATTTAAGCCAGCGGCAATCTTGGGACTAGATACCA 2397
Qy 578 agacctttatgcccagcaaaaaatgccattttttacagttctatcgagcggtggtgaggt 637
Db 2398 AAAGTGCTACAGCAGCAAAAAATCAACAATGGTTAAATCTACTCTTTTCGGGTACTC 2457
Qy 638 cgggtggttttgaagcactgttattcaactcaccacacggttaagagagaccataattc 697
Db 2458 ACAATGGCTTTTGTGCTCTTGTGATTTACATCCCGGTGATGGTAAGAAACGAAAGCTC 2517
Qy 698 attccgagggaataataataataataattcggcgctataaccggcgttgtaaaaaatcgct 757


```
LOCUS      APL17915      5200 bp      DNA      BCT      02-MAR-2000
DEFINITION Actinobacillus pleuropneumoniae rhoAP gene and ORF 2 & 3.
ACCESSION  Y17915
VERSION    Y17915.1  GI:4127359
KEYWORDS   rho gene; rho protein.
SOURCE     Actinobacillus pleuropneumoniae.
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Actinobacillus.
REFERENCE  1 (bases 1 to 5200)
            Tonpitak,W., Thiede,S., Oswald,W., Baltes,N. and Gerlach,G.F.
AUTHORS    Actinobacillus pleuropneumoniae iron transport: a set of exbBD
TITLE      genes is transcriptionally linked to the tbpB gene and required for
            utilization of transferrin-bound iron
JOURNAL    Infect. Immun. 68 (3), 1164-1170 (2000)
MEDLINE    20143720
REFERENCE  2 (bases 1 to 5200)
            Thiede,S.
AUTHORS    Direct Submission
TITLE      Submitted (11-AUG-1998) S. Thiede, Dept. of Microbiology and
JOURNAL    Infectious Diseases, School of Veterinary Medicine, Bischofsholer
            Damm 15, D-30173 Hannover, FRG
FEATURES   Location/Qualifiers
            1..5200
             /organism="Actinobacillus pleuropneumoniae"
             /strain="AP205"
             /serotype="7"
             /db_xref="taxon:715"
            271..1533
             /gene="rho"
            271..1533
             /gene="rho"
             /codon_start=1
             /transl_table=11
             /product="Rho protein"
             /protein_id="CAA76933.1"
             /db_xref="GI:4127360"
             /db_xref="SPTREMBL:O9ZAG7"
             /translation="MHLTQLKNTPPVSDLVALGEGOMLENLARLRKODIIFAILKHOH
             KSGEDIFGOGILFIPDGFCFLRSADSSVLGPDYVSPQIRRENFNLQTDGKIEGI
             RPPEGERFYALLKVLVDNDKPEVSRKILPENLPLHNSRLMERNGSTEDLTA
             RILDLAAPKGGQGLIVAPPKAGKTVLLQNIQSITHHPECELLVLLIDRPEREVT
             EMQRTVSGEIVASTDEPATRHVQVAEMVIEAKRSVEHKVDVILLDSITRLRAYN
             TVTPGKILSGVDANLHRPKRFAGARNVEGSLTIATALVDTGSKMDEVIFE
             EFKGTGMELHLSRKIAERRVFPPIEFNRSGTRKODLLMSPEHRNWLRLKVLNPD
             EVAAMEWLIDKLGVAKTNEEFELMKRS"
             complement(1581..3518)
             /note="ORF2"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="CAA76934.1"
             /db_xref="GI:4127361"
             /db_xref="SPTREMBL:O9ZAG6"
             /translation="MRKITPGGRIAPNPMKYQSGSWLFKLAHYHFSQHSVNAVFQI
             KQYDIQDMAQAKYSNLLPNNGKEPDPRKKEVEYSLFGRNFYDDPDSQAIGYNTLNP
             SLVRSAYWRARFEDEHHTKRLGFGYQVKSNDKGFIDLNLNLDDQDIRLLTLHS
             LSCSRYPYDKNCKITPKRDANATYRNITYKERKHQVOLSEKDLKIGKTOHLNLS
             MSYDRMSNRKRDQFERYLYVSKEDSEHPYIDVGNQDQGVYINSGYIKYPIRIIGK
             RQVIAHNNYCKFDMHYGKKKDFSDCSAHTTKQHYAIGLSDRIFFGDKLIDLGLVRY
             DYHRFKSDDDWNTGYTKNFSWNTGYTVRPREYIASSYRASNGFRVPSDFEMFGRVP
             GAADNYWRDLKKLTSEKALNQEVGVAFNGLGVLEVSYFDNRYRLIIVSAKEKG
             IDGIAGYNLQSLNLTGNVLNKLHWHGIWSALPDGLYNLAYNRILKRVNVDTKPNVY
             FVRSYVLDAIOERYVASLYGYDPNGKMGVNYLLTYSKAKKASELIGTYYTYGDVVK
             GYVNRQTRFWYIDVLYGYTFYFAKHTVRGVIYLNLTNRKYSTWEAVRQTAETNVIPIYN
             RGHASYAAPGRNVLSVEAKF"
             complement(4385..4996)
             /note="ORF3"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="CAA76935.1"
             /db_xref="GI:4127362"

BASE COUNT  1426 a 1191 c 948 g 1635 t
ORIGIN
Query Match 14.7%; Score 411; DB 2; Length 5200;
Best Local Similarity 51.3%; Pred. No. 3.7e-86;
Matches 1435; Conservative 0; Mismatches 1240; Indels 120; Gaps 16;
QY 72 atacggtgctgcgactgaaataaaaaatcgaagaaataaacgatatagctgtcttgcga 131
DB 4329 ATCCGCTTCTCTGATTTATCCGCACAAACTTATGCGAAGCAAAATGAACGGAATTAGA 4270
QY 132 tgaagttattgtacagagagccattatgtctcagcaacgtcaaaaagaagtaactgctt 191
DB 4269 TACCGTTACAGTCAAAACGAAACGGAGATACACCGCCGACAAAACGAGTAACAGGTTT 4210
QY 192 ggggaaagttagtgaataattatcacgaaatgagtaaaaaatacaattcttggattcgta 251
DB 4209 GGTTAAGTGACTATAAATCAGTAAGATCTTGATAAAGACGAGGTACAAAATATTCGTGA 4150
QY 252 tttaactcgtatgacctgtgatttcggtggtggacaaggtcgcggtgcaagttagtgg 311
DB 4149 TCTGACCCGTTATGATCCCGGTATTTCCGTTAGTTGACGAGGAGGAGGCCCAAGTAGCGG 4090
QY 312 ctatgcattcggaggtgtagataaaaacocgtgtcagcttacttctgtgtggttaccaca 371
DB 4089 CTATTCCGATTCGGGGGTAGATCGCAATCGTGTGGCATTACTAGTAGACGGGTACAGCA 4030
QY 372 aggcacagttatcatcagctaggttcagatgctaatggtggtgcaataatagagattga 431
DB 4029 AATTCAGTCTCTATAAAGTCTTCTCTGTTGCACAAAGCGGAGCAATTAATGAAATCGA 3970
QY 432 gtatgaaacattcgttcaattgagtttaagcaagagcagaagttctgcggaatagcttc 491
DB 3969 ATATGAANAATATTCGTTCGATTGAATTTAGTAAGGAGCCAGTTCACAGAATACGGTAG 3910
QY 492 tggtcgcatggtgtgctattggttttctgactaaagatgcgagcagattattataaga 551
DB 3909 CGGTGCTTATAGCGGTCGCGTCGCTCCGTCTACTATAAATCCCGACGATGTGATTAAAGA 3850
QY 552 ggggagcattggggttagatagtagaagacctttagccagcaaaaataagccattttt 611
DB 3849 AGGACAAAATTTGGGGATTAGACAGTAAACGGCGTATTTCGAGTAAATAAAGCAATTTAG 3790
QY 612 acagttatcgcagcggctggtgagcggtggtgttttgaagcactgtttattgcaactca 671
DB 3789 CGAATCGCTTAGTTTTCAGGTGCTGTTCCGCGGATTTGAAGCATTTGTCGAATATACGTA 3730
QY 672 ccgacacggttaagagacacaaatattcattccgagggcaataataataataataattccg 731
DB 3729 TCGAAAGGGCAAGAAACGGAAATTCATAAAGATGCGACT-AAGTAAGGCGAGTCCITCTA 3671
QY 732 gcgtataacgggtttgaaataatgcgtacactttaccataatccgcacagaatgctcct 791
DB 3670 CCGTTCGTGCCAAGAGAACTTAATTTGGGATAAAAAATCAAAAGAGATAAATCGTTAGTCTT 3611
QY 792 gggagttctcttttaattgtggaagatacttcccaacaattagattgtactcctgtgc 851
DB 3610 TGAAGACGAATGCCACCACGGGCAAGAGTAATCCCGGCTGTGTTCGCGGCTCCGATTGT 3551
QY 852 aagggttaagttagaacgcgcgataattccccagtgagaacaatttcccggaatatatacgctga 911
DB 3550 ACCAGGGGCG-----AAAGTGATAAAGAGATTCGACTAATTCGGAAGA 3509
QY 912 agagcgcaaacagcttgagcagattctctatcgcactgagcagctctcagcccaagaata 971
DB 3508 TTACACCAGGAGAGCGGCGCTTGTGCACCGAATCCTATG----AAATATCAGTCGGGTTCTT 3452
```

QY 972 tacccggttaagatcgcatcgaccccaacccttttagattacaagagtaattctgttttat 1031
Db 3451 GGTATTTAAATAGCTACCATTTCTACCGCAACATTCGGTAAACGGGTATTTGAAC 3392
QY 1032 gaagtttggtatcaacttaeactcgtctcattcttctggcgcaactctttagaagata--- 1087
Db 3391 AAACTAAGCAACGCTATGATATACAGGATATGCGACAAAGCAAAATATTCAAATTTATTAC 3332
QY 1088 caaaacacgctacgatatccgtgatatgcaaacgcccagcttactatacaaaagacgata 1147
Db 3331 CGAATAATAAAGGAGACCGGATCGTAATAAAGGATGGAATATTCGCTATTCGGTA 3272
QY 1148 ttaacttatcacttagaactatgtttatgaaggataataatttttagatggcttagtgt 1207
Db 3271 GAATTTCTATCCGGATGATCTCTCAACCAATAGGT-----TATGGAATACGC 3221
QY 1208 tcaagccaaggtacccttatgggttgcgtatagccatgtgaagtttttttgatgaacgtc 1267
Db 3220 TTAATCTAGTTTAGTGCAGAGTGCATTATTGGAGCAGAGGAGATTTTGTGACGAACGGC 3161
QY 1268 accaacaacgctgttaggattcacctataataataaacacagaga---ataatcgtggt 1324
Db 3160 ATACGAACACCCCTTAGGGTTTGGTATCAGTATAAGTCTGAAATGATAAAGGATTT 3101
QY 1325 tggatagcattaaactcagtcgsgataaaacagatattgaactatatagccgctacatc 1384
Db 3100 TTGATCGTCTAAATCTCAATTTGGATCAACAGATATTCGATTACTGACCAAGTTACACA 3041
QY 1385 gcttgcatttagcgtattccttctggttagataaaaaatccgcgcgactttggataaat 1444
Db 3040 GCCTAAGTTCTAGTAGTACCCTATTTTCGATAAAATTTGAAGATAACGCCGATAGC 2981
QY 1445 ctggtgtctatgtagaacgtagcgttaataattaccaagaagaacat-----cg 1493
Db 2980 GAGATGCGAATGTACTTATGACGGAATATTTATAGGAACGTCATAAACAAGTCCAAC 2921
QY 1494 tgcattcatttagaatttagaaagcgtataatgctggtcaaggcgtatttaacca--- 1550
Db 2920 TTTCTTTGAAAAGGATTTGGAATCGGGAAACGCAACACTTATTAAATTTATCCATGA 2861
QY 1551 -----aaccacaactgaatttaggttggcttggatcgattta 1591
Db 2860 GTTATGATCGTATCGCTCTCAACCAATAGCGAGATCAATTTGAACGGTTTATTAGTTT 2801
QY 1592 attcgcttatggatcatggtggtatgactgcccacat-----accaagggcgttata 1645
Db 2800 ATTCTAAGGAAGATGAATCGGAACATCCGTATTATGATGTTGGAAATCAAGATAAAGATG 2741
QY 1646 ccagctaccggttagaggcgttttagataatccatatattatcg----- 1691
Db 2740 TCGGTTATTACAATAGCGGTACTTATAAATACCCGAGAAATTCGGTAAAGACAAGCAA 2681
QY 1692 -ccgcatccacgcgattgaacgggtacttctgttgaataatacacgcggcgacatct 1750
Db 2680 TAGTCGCTCATATAATTATTGTAGTTTGTATGCAATATGGAACGCAAGATTTTA 2621
QY 1751 taaactgtgaaccgcgtataaattaaaggcgtatagccaattttgttagcttccgcgac--- 1807
Db 2620 GCGATTGTCGGACATATTTACCAAGGCGCAGCATTTACGATTTGGCTTATCGGATCGTA 2561
QY 1808 tagtgataacgagtagtgtgatttgggattaggggtcgtttttagatcaacatcgattta 1867
Db 2560 TTTCTTCGGTGATAAACTGGATCTAGGTTTGGGTGTACGGTATGATCATCTCGTTT 2501
QY 1868 aattgtatgcgttgacccttagccgaacttatccgaatttggcttggatgggtggga 1927
Db 2500 AATCGGATGATGTTGGTAAATACCGGTACATACAGAAATTTCTCTTGAATACGGGAG 2441
QY 1928 ttaacgttaaaccaacagagtttgtatcgttttcttctatcgcatttcaaacggttttagag 1987
Db 2440 TGACGGTTCGACCGAGAGATAATATCGCTTATCTTACCAGGCTTCCAACGGTTTTAGAG 2381
QY 1988 tgcctgcattctatgaactttatgggttaaacgctgatcatattggccttaaaagataaagaat 2047

Db 2380 TTCGGTCATTTGATGAATGTTTCGGTCTCGAGTTC---CCGGTCCGCAGATAACAAT 2324
QY 2048 atgtgcaacgcgcgcaacgtagccaccagttagagcagaaaaaatcgcataatcatgaga 2107
Db 2323 ATTGGCGTAGAGACTTTACTTAAGAAAAAATAACTTCGGAAGAACACTTAACCAAGAAG 2264
QY 2108 ttggagtagctttaaaaggtcaatttgggttaacttgatgtgagctatttccgtaaacat 2167
Db 2263 TCGGAGTGGCGTTTAAACGGCAATTTAGGAGTATTGGAAGTCAGTTATTTTGTGATAACCGTT 2204
QY 2168 ataaaaaatgatgtgcgacgcatgttaaagaataatacaaaaaatcacactgtttctata 2227
Db 2203 ATCGAGACTTGATGTGTCAGTGGCAAGAAAAAGGGATTGACGGAATAGCGGGTTTCTATA 2144
QY 2228 ataccataaatattcaagatgtagcaactaaacgggataaatttagtcgctaaatttgact 2287
Db 2143 ATCTCCAATCGATTAAAT-----CTTAATGGGGTAAACCTATTGGGTAAACTCAAT 2093
QY 2288 tacacggtattttatctatgctgcagatgggttttttattcatcagttgcttataaacgtg 2347
Db 2092 GGCACGGTATTTGGTCGGCTTTACCGGACGGCTTATATTCAAATCTTGCTTATAACCGTA 2033
QY 2348 taaaagtaaaagcggaaactaaacgcgactcaagactcgatagcgtaaacgcatctatcc 2407
Db 2032 TTAAGTGAATAACGTCGA---TACGAAGCCGAATTAACGTGTTGTACGTAGTCGGTAT 1976
QY 2408 tagatgcgattcagccagcacgctatgtctgttgattcggttcgctacatcccccagaaga 2467
Db 1975 TGGATCGGATACAGCCGGAACGCTATGTAGCCAGCTTAGGTTATGATGAGCCGAACGTA 1916
QY 2468 aatggggaattggcattactaccacctattctaaagccaaacacgcgcatgaggtggcag 2527
Db 1915 AGTGGGGTGTGAATTAATTTCTAACTTACTTAAACGGAAGAAAGCCAGTGAATTAATCG 1856
QY 2528 gcaacgctcatcagcgatcacatcgctgtttagttggtg-----caaacgtaccggtt 2580
Db 1855 GTACGACGACTTACTATATCTGGAGATCTGGTAAAGGTGTGACGAATCAACGTACTCGAC 1796
QY 2581 cttggtacacccatgatattaccgg---ttacatcaattataaaaaactcacacttaccgtg 2637
Db 1795 CTGTGTATATTTATGATGTTATTGGTTATTACACCTTTGCGAACAATTTTACCGTACGAG 1736
QY 2638 gaggaatttataatgtgactaatcgttaaatattccacttgggaatcagtcgccaatccg 2697
Db 1735 GTGGTATTATACCTAACCAATCGTAATACAGTACAGTACGTTGGGAAGCGGTTCGCCAACTG 1676
QY 2698 gtgtgaatgcagtaaaacgaacgcgggttagcaattcacactgatttggcgtcccgggga 2757
Db 1675 CGAAACCAATGTGATTCGGATTTAATTCGAGGTCTATCGGAGTTATGCCGACCCGGCA 1616
QY 2758 gaaatttcagtttagcatttgaattgaagtttag 2792
Db 1615 GAAACTATGTACTAAGTGTAGAGCTAAATTTCTAG 1581

RESULT 8

U32780/c

LOCUS

DEFINITION

U32780 L42023

ACCESSION

U32780.1

VERSION

U32780.1

KEYWORDS

SOURCE

ORGANISM

Haemophilus influenzae Rd.

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

Haemophilus.

REFERENCE

1 (bases 1 to 10205)

AUTHORS

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,

Kirkness, E.F., Kerevage, A.R., Buit, C.J., Tomb, J., Dougherty, B.A.,

Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,

Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,

Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,

U32780

10205 bp

DNA

BCT

29-MAY-1998

Haemophilus influenzae Rd section 95 of 163 of the complete genome.

U32780.1

GI:1574020

Haemophilus influenzae Rd.

Haemophilus influenzae Rd

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

Haemophilus.

1 (bases 1 to 10205)

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,

Kirkness, E.F., Kerevage, A.R., Buit, C.J., Tomb, J., Dougherty, B.A.,

Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,

Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,

Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,

Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C.
Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd

Science 269 (5223), 496-512 (1995)
93350630

JOURNAL MEDLINE
REFERENCE

AUTHORS

2 (bases 1 to 10205)
Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S., Borodovsky, M., Rudd, K.E. and Koonin, E.V.
Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*
Curr. Biol. 6 (3), 279-291 (1996)

JOURNAL MEDLINE
REFERENCE

AUTHORS

3 (bases 1 to 10205)
White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
Direct Submission

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REMARK
The whole genome was shifted by 588 nucleotides for a new start
On Sep 30, 1996 this sequence version replaced gi:1221734.
Location/Qualifiers
1. .10205

JOURNAL TITLE
Submitted

JOURNAL TITLE
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL TITLE
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

/db_xref="GI:1574022"
/translation="MDFSSISRENLLKPLQOVCVGLNRPNPVNLNNVLQIEDYRLTI
TGLDVEVLSSTQQLSSSENGFTTIPAKFLDICTLSDSDSEITVTFQDRLVQSG
RSFETATPAEYPNLTQWSEVDFELQNTLRLRIEATQFSMAQDARYFLNGMRF
ETENLLRVADVCHRLVACTISLEQLQNSVILPRKGVLELRLLETNDEPARLOI
GTNNLRVHLKNTVTSKLIDGRFPYRRLPRNATKIVGNNWMLKQAFARASILNE
RARSVLTKENOLKTASNTHEEAEIIVDNYNGELEVGFNVYIILDVNLALCN
QVMCLTDAFSSCIENCEDSSCEYYVIMPLRL"

gene

complement(2345..3709)

CDS
complement(2345..3709)

/gene="HI0993"

/note="similar to GB:L10328 SP:P03004 GB:J01602 PID:145760
PID:290550 percent identity: 62.36; identified by sequence
similarity; putative"

/codon_start=1

/transl_table=11

/product="chromosomal replication initiator protein
(dnaA)"

/protein_id="AAC22655.1"

/db_xref="GI:1574023"

/translation="MNLNWSCLLQLOQVVSASLSTWLRPLQADVADNHIIVYA
SNMFVGVETHYLAQIQICOTLAQNPRLISLKEGVAPKPIVESTNLSRSEA
VDFQESSASVKEFSLNTHLPDFEGKSNQARAVGKLAQAFGEPSANFFLYG
GTGLGTHLLHAIGNLADKPNARVLYTHANNMOHVMKAVDRKMDQFKFYSLD
ALLVDIOQFAKEKTOEFFFHFNLSFTQCIILTSDRYKEIEKIERLKSFGW
GTTALEPDLTRVAILLKAEEHNNLPEVAFIAQRLTNRELSGALNRKVM
QDFKGDIDIDFVRDLKDLALQERLVNIENIKVVAETIRIKVSDLSKSRARSV
TRFQIAMAALAKELNLSLPEIGRAFDRDHTTLLNACREVPKFREQDNTQEDWANLIR
TLISA"

gene

complement(3901..6639)

CDS
complement(3901..6639)

/gene="HI0994"

/note="similar to GB:215129 SP:Q06987 GB:S64486 GB:X75168
PID:297044 percent identity: 51.19; identified by sequence
similarity; putative"

/codon_start=1

/transl_table=11

/product="transferrin-binding protein 1 precursor (tbp1)"

/protein_id="AAC22656.1"

/db_xref="GI:1574024"

/translation="MTKPEFRLSIISCLLSYKVAETQSIKDKAEISSEVDOST
EDSELETISVTAEKVRDRKDNVTGLKIITSEISREQVNLNIRDLTRYDGLISVVE
QGRGSSGYSIRGMDNRVALLVDGLPOTQSVVQSPVARSYSCTGAINLEIYEN
KAVEISKGSSEYGNALAGSVTQSKASADILGDSKNGITKNAYSKNKGFTHS
LAVAKQGFEGLAITORSLETQVHKDALGVQSNRLIAKPENQSAIFYNDECP
KYDCEIPSAKPPAILSTKRTSVSDYTGANRIPNPKMYESQSLFGGYSFQKH
YIGGIFETQQRFDINDMTPPAYLPTPEKGLANRPFPYKQDYGAYQHTDGRGVYA
SGLYFDEHHRKORGVIEYIENKAGIIDKAVLSANQONILDSYMRHTHSLYENP
SKNCRPTLDKPYSYRSDENVYKEHNMQLNLEKKIQONWLTHQIVFNLGDDFTSA
LQHKDYLTRVATADSDTKYKTRNGRLREPYIPKPKPFAGEDHCNQSSEN
YRDKVRLLIKGNYFAARNMALGVLDGLGIRYDVSRTKANESTISVGRKPNFSW
NTGIVIKPTWELDSLSTGRPNFSEAFEMGYRGKNDEYVVKFETSRGNQFEG
LALKDGFKEIHSFNAYNLIAFAELSKNGTGTNGYNAQNAKLVGNVITAOI
DFNLGKRPYGYWATVAVNRVYKDKQINAGLASVSLFDIAIOPRSYIIGLDHI
SNWGNATFTOSKAKSONELLGKALGNNRSDVSTRKLTARHILDSVGYMANKN
IMLRGILNLFNIRYIVTAVQTAQAGVNHQNVGNTRYAASGRNYTLLEMF"

gene

complement(6667..8544)

CDS
complement(6667..8544)

/gene="HI0995"

/note="similar to PID:1177569 percent identity: 39.97;
identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="transferrin-binding protein 2 precursor (tbp2)"

/protein_id="AAC22657.1"

/db_xref="GI:1574025"

/translation="MKSVPILSGLSLFLSACSGGGGSDVDVSNPSSKPRYQDDT
SNORKSNLEKLSIPSLGGMKLVAQNLGRPSLINEDGYIIFSLSKIEDDFKEN
QSQTEPTIGSIDEPSTPNHGGQYVYSGLYIQSWRNFSNGFYSGYGYAYFG
KQATATLPVNGEATYKGTWSTFATATGRKGYSLFSNSSSQGYSRRSAISEDLENDQ

```

NGETGLISQFSADFGTKKLKGLFYTKRKTNNQNYEKKKLYDIDANIYSNRFGRKVK
PTKDEEHPFTREGTLEGGFYGPNGEELGGKFLAGDKKVFVGSFAKEEYKOKTLP
KETLIDGKLTFTSTKPDATSTTANAKDTAENFTKDISSEGEADYLLIDNP
VPLLPTEGSGDPATSKHVEVRDKTYKVEACCKNLKYVFGMYEDENKKNKNETEQY
HOLLGLRTASSKIPGTGNVYKYSWFGYSIDGETSYSTGDKRDKNAVAFEDVNF
EKLKGLSRADQNPVFSEIANFKNGNAFTGTATAKDLVIDGKNSQTKNPTINITT
KVNGAFYGNASBLGSGFYTYNGKNPTDKNSPTASSPSNSEKARAAVVFAGKKQVETNN
K"
6642..9082
/gene="HI0996"
CDS
6642..9082
/feature="H. influenzae predicted coding region HI0996"
/feature_id="AAC22658.1"
/db_xref="GI:1574027"
/translacion="MNGVKQLFLLSLIGLSLNVAAVEARPKNDTLNTIOSAELK
TSFSFSPKKEIPNRHIIISLSKSLAHRLVLRGLIPALYQNNQVQVLLPLYKQF
PQDNFLLTWAKAIEAREGDLTQSIAYIRELFARDASLLPLRY"
9398..10087
/gene="HI0997"
CDS
9398..10087
/feature="H. influenzae predicted coding region HI0997"
/feature_id="AAC22659.1"
/db_xref="GI:1574026"
/translacion="MNGNGKYWDNKKYNENATVRIGGGLGYOTAVEVSLPFPQEKR
WYAGSSGNTMKQYADKLGIRLENWLSKTDQISTALEYGESRYKIRKHLGDNIFY
VSTLFLPKSTQFVFGMDFHRENTQALDNOYQKTLRLGWQDMFYGISRLTFSY
ANRVYREKDLIGIQKNREYTTITLHRNHPMLTLPKLSWDYQKSTNSHAFYRDK
NRYLIGIKIF"
BASE COUNT      3008 a 1866 g 1840 g 3491 t

Query Match      12.4; Score 345.2; DB 2; Length 10205;
Best Local Similarity 51.2; Pred. No. 1.1e-70;
Matches 1370; Conservative 0; Mismatches 1123; Indels 181; Gaps 17;

Qy 173 aaacgaagtaactggtgggaagtagtgaataattatcacgaataatgataaaatc 232
Db 6449 ATAATGAGTAACTGGACTTGGCAAAATTAATCAAAACTATGTAAGTATCAGCCGGAAC 6390

Qy 233 aaattctgtgattcgtgatttaactcgctatgacctggtatttcggtggtggaacaag 292
Db 6389 AAGTATTAAATATTCGTATCAACAGCTATGATCAGGGATTTTCAGTTGTACACAAG 6330

Qy 293 gtcgcggtgcaagtagtgctatgccattgaggtgtagatataaaacccgtgctagcttac 352
Db 6329 GTCGCGGTGCAAGTCTCGGATATTCATTCGTGTGTGACAGAAATAGAGTTGCTTTAT 6270

Qy 353 ttgtgtatgggtaccacaagcgacaggttatcatatcagctagggtc----- 398
Db 6269 TGGTAGATGTTTACCTCAACAGCAATCTTATGTAGTCAAGCCCTTTAGTTCCTCGTT 6210

Qy 399 ----agatgctcaatgggtggtcaatgaattgagattgagatgaatactgctcaattg 454
Db 6209 CAGGATATTCGGCAGCTGGTGCATTAATGAATTTGAATATGAATTTGAAGCCCGTCG 6150

Qy 455 agttaagcaaaaggagcaagttctgcgaataatggctctggtgcgcatggtggtgctattg 514
Db 6149 AAATAAGCAAGGGGGAGTCTTCTGAGTATGGAATGGAGCATTAGTCTGTTCTGTAA 6090

Qy 515 gttttcgtactaaagatgctgcagatattattaaagggggcagcattggggttagata 574
Db 6089 CATTCAAGCAAAATCCGACGCCCATCTCTTGAAGGAGAGACAAATCATGCGGGAATCAAA 6030

Qy 575 gtaagacctcttattgccagcaaaaatagccatttttttacagtctatcgacgagcgtggtg 634
```

```

Db 6029 CTAAAAATGCTTATTCAAGCAAAAAATAAGGCTTTACCCATTCTTTAGCTGTAGCAGGAA 5970
Qy 635 agcgcgggtggttttgaagcaactgttattgcaactcacgcagcagcggtaagaagacccaaaa 694
Db 5969 AACAAAGGTGGATTGAAGGACTTTGCTATTATACACTCAACGAAATTCATTTGAACCAAG 5910
Qy 695 ttcatccagaggaataataataataattccggtataaccgcggttttgaaatc 754
Db 5909 TCCATAAAGATGCTATAAAGGGGTGCAAAAGTTTATATGCTTTAATCGCC----- 5858
Qy 755 gctacgacttttaacccaaattccgcacagaatgctcctggaggatcctctttaaattgtg 814
Db 5859 -----AAACCAGAGAAATCAATTCGCATACTTIG-----TGATGG 5826
Qy 815 aagatacttgcacaaacattagattgtactcctcgtgcaaggggttaagttgaacgcgata 874
Db 5825 AAGATGAGTGTCCA-----AAGGGTTATGACGAGTGTATACC 5789
Qy 875 atttccagtgagaacattttccggaataatacgcctgaagagcgaacacagcttgagcaga 934
Db 5788 TTCAGC-----CAACCCACTGCGGATTT 5766
Qy 935 ttcttctatgcactgagcagctctcagcccaagaataataccggtataaagatcgcattgca 994
Db 5765 TATCCACCAAAAAAGAACCGTAAGCGTTTCAGATTATACGGGGCTAACCGTATCAAAAC 5706
Qy 995 caaaccttttagattacaagagtaattctgtttttttagaagtttggctatcaactcaact 1054
Db 5705 CTAATCCAATGAATATGAAGCCAGCTTTGGTTTTTAAGAGGAGGTATCATTTTTCTG 5646
Qy 1055 cgtctcatctctggcgaactcttagaagatacaaaaacacgcacgtacatccgata 1114
Db 5645 AACAACTATATATGGTGGTATTTTGAATTCACACAAAAAATTTGATATCCGIGATA 5586
Qy 1115 tgcaaacccagcagctactactacaaaaagacgatatctaacttatcacttaggaactatgttt 1174
Db 5585 TGACATTTCCCGCTTATTTAAAGCCCAACAGAAAAAGGGGATTTAGCAAAATCGCCCTTTT 5526
Qy 1175 atgaaggggataatttttagatggcttagtggttcaagccaagagatcccttaagggtgc 1234
Db 5525 ATCCAAAGCAAGAT-----TATGGCGCATATCAACATATTGAGGATGGCGAGGCGTTA 5472
Qy 1235 gctatgcccctggaagttttttagtgaaacgtcaccaacaacgcgtcttaggtatccact 1294
Db 5471 AATATGCAAGTGGCTTTATTCGATGAACACCATAGAAACACGCGTGTAGTATTGAAT 5412
Qy 1295 ataaataaaacacagagaataa---tcgctggttggtatagcattaaactcagtcgagata 1351
Db 5411 ATATTTACGAAAAATAGAACAAAGCGGGCATCATTTGACAAAGCAGTGTAAAGTCTAATC 5352
Qy 1352 acaagatatgtaactatagccgcgtacatcgcttgctgattgtagcgattatcctctg 1411
Db 5351 AACAAACATCATATCTTGACAGTTATATGCGACATAGCATTCGAGTCTTTTCTCTAATC 5292
Qy 1412 tagataaaattccgcgcacatttggaataaacttggtctatgtatcgaactcagcgtata 1471
Db 5291 CAGTAAAGATTCCGCCCAACACTTGTATAACCTTATTCTACTACTCTGTTCTGATAGAA 5232
Qy 1472 ataattaccaagaaaagcagctgctcatttatttagaatttgataaa-----gcgctaa 1525
Db 5231 ATGTTTATAAAGAAAAACATAATATGTTGCAATTTGAATTTAGAGAAAAAATAATCAACAAA 5172
Qy 1526 atgctggtgcaagcgtattttaaaccacaccccaaaaactgaatttagggttgggcttgatc 1585
Db 5171 ATTGGCTTACTCATCAAAATTTGCTTCAATCTTGGTTTTGATGACTTTTACATTCGCCACTTC 5112
Qy 1586 gatttaattcgtctatgatacatgagggtatgactgcccataataccaaagggcgttata 1645
Db 5111 AGCATAAAGATTAATTAACTCGACGTTTATTCGCTACGGCAGATAGTATTTTCAGATAAAA 5052
Qy 1646 ccaagctaccgcggttagaggcggttta--gataatccatataatttatcgccgagctccacgc 1704
```


Db	4106	GTITTTGCTGAATGATGTTGGCTGATGGTGGCAATAATA-----	4147
Qy	2054	aacgcgcgaacgtagccaccagtttagagccagagaaatcgcactaatcatgagattggag	2113
Db	4148	---GCGAGGTTTATGTAGTAAATTAAGCCTGAACATCTCGTAACCAAGATTGTGTC	4204
Qy	2114	ttagctttaagagtcgaattggtacccttgatgtagctatttcogtataaactataaaa	2173
Db	4205	TGCGCTTAAAGGGGATTGTGTAATAATGAGATCAGTCATTTAGTAAATGCTATCGAA	4264
Qy	2174	atatatgtagcagacgcatgtaaa---agaataatacaaaatcacactgtttctataact	2230
Db	4265	ATCTTATCGCCTTTGCTGAAGAACCTTATAAATAATGGAACCTGGAAAGGCCAATTATGGAT	4324
Qy	2231	accataatatcaagatagcactaaacgggataaaatttagtcgtctaaatttgacttac	2290
Db	4325	ATCATAATGCACAAAATGCAAAATAGTTGGGTAAATAATAACTGCGCAATTAGATTTA	4384
Qy	2291	acggtatttatctatgctgcagatggtttttattctatcctcggtgcttataacogtcaa	2350
Db	4385	ATGGTTTATGGAAGAGTATTCCTACGGTTGGTATGCAACATTTGCTTATAACCGAGTAA	4444
Qy	2351	aagtaaaagcgggaacactaacggactcaagactcgatagcgtaaacgatcctattctag	2410
Db	4445	AAGTTAAAGATC---AAAAATCAATGCTGGTTGGCCTCGTAAGCAGTTATTATTATGG	4501
Qy	2411	atgcgattcagcgcagcagctatgctgtgattcggtacgataccaccagaagaataat	2470
Db	4502	ATGCCATTACGCCCCAGCGGTTATATCATTTGGTTAGGCTATGATCATCCCAAGTAATCTT	4561
Qy	2471	ggggaaattgycattactaccacctattctaagccaaaacgcgcgatgagtgccaggca	2530
Db	4562	GGGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGCTAGGAA	4621
Qy	2531	cagtcacatcagcagataatcgc-----gttgattaggtaggcgaactgacccggttctt	2583
Db	4622	ACGTGCAATGGGTAAACAATTCAGGGATGTAAATAACAAGAAAACTTACTCGGGCAT	4681
Qy	2584	ggtacaccatgatattaccgg---ttacatcaatataaaaactacaccttagctggag	2640
Db	4682	GGCATATCTTAGATGATACGGGTATATACATGGCGCAATTAATAATATTATGCTTCGATTAG	4741
Qy	2641	gaattataatgactaatcgttaaatattccacttgggaatcagtgcccaatcccggtg	2700
Db	4742	GGATATAAATTTATTCAACTATCGTATGTTACTTGGGAAGCGGTGGCTCAAAACAGCAC	4801
Qy	2701	tgaatgcagtaaaccaagaccgg---gtagcaattacactcgatttggcgctccgggga	2757
Db	4802	AAGTGGCGGTCAATCAACATCAAAATGTTGGTAGCTATACTCGTACGACGATCAGGAC	4861
Qy	2758	gaattttcagtttagcatttgaatgaagtttta	2791
Db	4862	GAACATACTACCTTAACATTAGAAATGAATTCATA	4895
RESULT 10			
I79990			
LOCUS I79990 5099 bp DNA PAT 10-JUN-1998			
DEFINITION Sequence 4 from patent US 5708149.			
ACCESSION I79990			
VERSION I79990.1 GI:3208280			
KEYWORDS			
SOURCE Unknown.			
ORGANISM Unclassified.			
REFERENCE 1 (bases 1 to 5099)			
AUTHORS Loomore,S., Harkness,R., Schryvers,A., Chong,P., Gray-Owen,S., Yang,Y., Mordin,A. and Klein,M.			
TITLE Method for producing purified recombinant Haemophilus Influenzae transferrin binding proteins			
JOURNAL Patent: US 5708149-A 4 13-JAN-1998;			
FEATURES Location/Qualifiers			

source	1. 5099	/organism="unknown"	
BASE COUNT	1834 a	849 c	960 g 1456 t
ORIGIN			
Query Match	12.2%;	Score 340.8;	DB 5; Length 5099;
Best Local Similarity	50.4%;	Pred. No. 1.1e-69;	
Matches 1347;	Conservative	0; Mismatches 1152;	Indels 175; Gaps 14;
Qy	173	aaaacgaagtagcgttggggaagtagtaaaattatcacgaaatgataaaatc	232
Db	2342	ATAATGAAGTAACCTGGACCTTGGCAAAATATCAAAACTAGTGAAGTATACGCCGAGAA	2401
Qy	233	aaattcttggtattcgatttaactcgctatgacctggtatttcggtggtggaacaag	292
Db	2402	AAGTATTAAATATTCGTGATCTACACGCTATGATCAGGCATTTTCAGTTGTGAACAAAG	2461
Qy	293	gtcgcggtgcgaagtagtggctatgccattcgaggtgtagataaaacccgtgtcagcttac	352
Db	2462	CCCGTGGTGAAGTTCGTGATATTCTCTGTTGTTATGTCGACAGAAATAGAGTTGCTTTAT	2521
Qy	353	ttgttgatggctaccacaagcgacagttatacagctaggttc-----	398
Db	2522	TAGTAGATGGTTTACCTCAAAACGCAATCTTATGTAGTGCRAAGCCCTTTAGTGTCTGTT	2581
Qy	399	---agatgctcaatggtggtgcaatttaagattgagttgagtaaaacattcgttcaattg	454
Db	2582	CAGGATATTCGGGCACATGGTGCATTTAATGAATTTGAATATGAATAATGAAGGCCGTCG	2641
Qy	455	agttaacaaaggagcgaagttctcgcgaatatgctcctggtgcgcattggtggtctattg	514
Db	2642	AAATAAGCAAGGGGGGAGTTCTTCTGAGTATGTAATGGAGCAGTACTAGTGGTCTGTAA	2701
Qy	515	gtttctgactaaagatgcgcaggatattataaagggggcagcattggggccttagata	574
Db	2702	CATTTCAAAGCAATACGACAGCCGATATCTTAGAAGGAGACAAATCATGGGGAATTCAAA	2761
Qy	575	gtaagaccttattgcagcaaaaaatagccattttttacagctctatcgacgcgctgtg	634
Db	2762	CTAAATATGCTTATTCAGCAAAATAAAGGCTTTACCATTCTTTAGCTGTAGCTGGAA	2821
Qy	635	aggcggtggttttgaagcactgtttattgcaactcaccgcacgcggttaaaagagaccaaaa	694
Db	2822	AACAAGGGGATTTCAGCGGTGCGCATTTATCTCAACGAAATTCATTAATCAAAACCCCAAG	2881
Qy	695	ttaattcagaggcaataataataaaacataatttcggcgatataacccggctttgaaatc	754
Db	2882	TCCATAAAGATGCATTTAAAAGGCGGTACAAAAGTTATCATCGATTAAATCGCC-----	2936
Qy	755	gtcacgactttacccaaattccgcacagaatgctccctggagatctccttttaattgtg	814
Db	2937	AGAGGATCAATCTGCATACTTTGTGTGATGCAAGATGATGTCCAAAGCCAGATGATTATAA	2996
Qy	815	aagatacttggcccaacattagattgtactcctcgtgcaagggttaagtgaacgcgata	874
Db	2997	CAGTTGTTTACCTTTC-----	3012
Qy	875	atttccacgtgagaaacatttccgggaatatacgcctgaagagcgcaaacagcttgagcaga	934
Db	3013	-----GCCAAACGACCTGCGATTT	3031
Qy	935	ttccttatcgactgagcagctctcagcccaagaatatacgggtaagagatcgcttgcac	994
Db	3032	TATCTCCCAAGAGAAACCGTAAGCGTTTCAGATTATACGGGGGCTTAACCGTATCAAAC	3091
Qy	995	caaaccttttagattacaagagtaattctgttttttgaagtttggcttaccactcaact	1054
Db	3092	CTAATCAATGAATAATGAAGCCAGCTTGGTTTTTAAAGGAGGGGTATCATTTTCTG	3151
Qy	1055	cgctcattatcttggcgcaatctttagaagatacaaaaaacgcgtacagatccgtgata	1114
Db	3152	AACAACATTATATTGGTGGTATTTTGAATTTCACACACAAAAAATTTGATATCGTGTATA	3211

Qy 1115 tgcaaacgcagcttactatatacaaaagacgatattataacttatacttaggaactatgttt 1174
Db 3212 TGCAATTTCCCGCTATTATTAGATCAACAGAAAAACGGGATGATAGCAGTGGCTCTTTT 3271
Qy 1175 atgaaggaggaataattatttagatggcttagtggttcaagccaaggtatcccttatgggtgc 1234
Db 3272 ATCCAAACAGATATATGGTGCATATCAACGTATTGAGGATGGCGG-----AGCGGTTA 3325
Qy 1235 gctatagccatgtgaagttttttgatgaacgtaccccaaaacgcgtctgttaggtatccct 1294
Db 3326 ACTATGCAAGTGGGCTTATTTCGATGAACACCATAGAAAAACAGCGTGTAGGTATTGAAT 3385
Qy 1295 ataaataataaacagagaataaatcgcgtgtgttgtagcattaaa---ctcagtcgggata 1351
Db 3386 ATATTTTACGAAAAAAGAAACAGCGGCATCATTTGACAAAAAGCATGTGTAAAGTCTTAATC 3445
Qy 1352 aacaagatatgaactatatacgccggtcacatcgcttgcattgttagcattatcctgtgg 1411
Db 3446 AACAAACATCATCTTGCACAGTTATATGCAACATACGCATTGCAGCTTTATCCTTAATC 3505
Qy 1412 tagataaaaaattgcgcgcgactttggataaaattcttggctatgtatgcgaactgagcgta 1471
Db 3506 CAAGTAAGAATTGCGGCCAACACACGCTGATAAACCTTATTCTACTATTCATCTCTGATAGA 3565
Qy 1472 ataattaccagaagaacatcgtgtcattcattcatttagaatttgataaagcgcgtaaatgctg 1531
Db 3566 ATGTTTATAAGAAAAACATATATGTTGCAATTGAAATTTATAGAAAAAAATTTCAACAAA 3625
Qy 1532 gtcaaggcgtatttaaccaccccaaacactgaatttaggg-----ttggccttg 1582
Db 3626 ATTGGCTTACTCATCAAAATGCTTCATCTTGTTTGTGATGACTTTACTTCACGGCTTC 3685
Qy 1583 atcgattttaactcgtttagatcatcgtgggatactgactgcccaataataccaaaagcgggtt 1642
Db 3686 AGCATAAAGATTATTTAACTCGACGTGTACCGCTACGCAAGAGTATTTCAGAGAAAG 3745
Qy 1643 ataccagctacgcggttagagggggttagataatccataatttatccgcgcgatccac 1702
Db 3746 CTAATGAACCAAGAGAATGTTTACAAAAAACACCTTACTTTATACCAAAACCAACAG 3805
Qy 1703 gcagttatgaacggtatcttctgtgtaataacacg-----cggcacatcttaact 1756
Db 3806 TAGGTTTGTAGTACACAGATCATGTGATATATAAGGTAACTCTCTTAATTACAGAGACT 3865
Qy 1757 gtgaaccgcgtataaaagcgatagccattttttagcttcgc---gatctagtga 1813
Db 3866 GTAAAGTGGGTTAATTAAAGGGAATAATTATTTCGCAGCAGCAATAATATGGCAT 3925
Qy 1814 taagcgagtagtggtatttggttaggggtgcgttttgatcaacatcgatttaaatctg 1873
Db 3926 TAGGGAATACGTTGATTAGGTTTAGGTATTCGGGTATGACGTATCTCGCACAAAAGTA 3985
Qy 1874 atgacctggaacacttagccgaaactatcgaaactgtcttggaatggtgggattacgc 1933
Db 3986 ATGAATCAACTATTAGTGTGTAATTTAAAAATTTCTCTTGGAACTACTGGTATTGTCA 4045
Qy 1934 ttaaacacacagagttgtatcgcttcttctatcgcatcttcaaaaggttttagagtgctg 1993
Db 4046 TAAACCAACGGAATGGCTGTATCTTCTTATCGCTTTCTACTGGATTAGAAATCCTA 4105
Qy 1994 cattctatgaactttatggtaaacgtagatcatattgggcttaagataaacgaatatgtgc 2053
Db 4106 GTTTTGTCTGAATGATGTGTTGGCGGTATGTTGGCAATAATA----- 4147
Qy 2054 aacgcgcgaacgtagccacacgtagccagagaataatcgactaatcatgagattggag 2113
Db 4148 ---GGCAGGTTTATGTAGTAAATTTAGCCCTGAACATCTCGTAACCAAGATTGGTC 4204
Qy 2114 ttgattttaaagggtcaatttggttacccttgatgtgagcattttccgttaataactataaaa 2173
Db 4205 TCGCTCTAAAAGGGGATTTGGTAATATTGAGATCAGTCAATTTTAGTAAATGCTTATCGAA 4264

Qy 2174 atagtattgcagcagcatgtaaa---agaataatacaaaaaacacactgtttctataact 2230
Db 4265 ATCTTATCGCTTTGCTGAAGAACTTAATAAAAAATGAAGTGAAGGCCAATATATGGAT 4324
Qy 2231 accataatttcaagattagcactaaacgggataaatttagtcgcgtataaatttagcttac 2290
Db 4325 ATCATATGCAAAAAATGCAAAATTAGTTGGCGTAAATAATAACTGCGCAATTAGATTTTA 4384
Qy 2291 acggtatttatctatctgctccagatgggtttttatttattcattcagttgcttataaacggttaa 2350
Db 4385 ATGGTTTATGAAACGTAATTCCTTACGTTGGTATGCAACATTTGCTTATAACCGAGTAA 4444
Qy 2351 aagtataagcgggaaactaacgcactcaagactcagatcgatagcgtataaacgactcctattctag 2410
Db 4445 AAGTTAAAGATC---AAAAATCAATGCTGTTTGGCTCCGTAAGCAGTTATTATTATG 4501
Qy 2411 atgcgattcagccagcgcgtatgtgcttgcgtacgtatcggtcgcgtaccccgagaagaaaaa 2470
Db 4502 ATGCCATTACGCCAGCCGCTTATATCATTTGTTTAGCTATGATCATCCAAGTAATACTT 4561
Qy 2471 ggggaattggtcattactaccacattcttaagccaaaaacgcgcgatgggtgagcgca 2530
Db 4562 GGGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTTGCTAGGAA 4621
Qy 2531 cagctcatcagcatatcgcg-----gttgatttaggtggcaaacactgacgcgttctt 2583
Db 4622 AACGTGATTTGGGTAAACAATTAAGGGATGTAAATCAACAAGAAAACTTACTCGGGCAT 4681
Qy 2584 ggtacacccctgatcattaccgg---ttacatcaattataaaaaactacaccttacctgtgg 2640
Db 4682 GGCATATCTTAGATGATATCGGTTATTACATGGCGAATAAAAAATATTATGCTTCGATTAG 4741
Qy 2641 gaattataatgtgactaatcgttaaatattccacttgggaatcagtgcgcaatcccggtg 2700
Db 4742 GGATATATAATTTATTCAACTATCGTATGTTACTTGGGAAGCGGTGCTCAAAACAGCAC 4801
Qy 2701 tgaatgcagtaaaacaaagaccgg---ggtagcaattacactcgatttggcgtccgggga 2757
Db 4802 AAGTGGCGGTCAATCAACATCAAAATGTTGGTAGCTATATCTCGCTACGAGCATCAGGAC 4861
Qy 2758 gaaatttcagtttagcatttgaaatgaagttaa 2791
Db 4862 GAAACTATATCTTAACATTAGAAATGAATCTTA 4895
RESULT 11
HIU15058
LOCUS
DEFINITION
Haemophilus influenzae SB33 transferrin binding protein 1 (tbpl)
gene complete cds.
ACCESSION
U15058
VERSION
U15058.1
KEYWORDS
GT:1223954
SOURCE
Haemophilus influenzae.
ORGANISM
Haemophilus influenzae
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE
1 (bases 1 to 5144)
AUTHORS
Loosmore,S.M., Yang,Y.P., Coleman,D.C., Shortreed,J.M.,
England,D.M., Harkness,R.E., Chong,P.S. and Klein,M.H.
Cloning and expression of the Haemophilus influenzae transferrin
receptor genes
Mol. Microbiol. 19 (3), 575-586 (1996)
JOURNAL
MEDLINE
REFERENCE
2 (bases 1 to 5144)
AUTHORS
Loosmore,S.M.
TITLE
Direct Submission
JOURNAL
SUBMITTED (23-SEP-1994) Sheena M. Loosmore, Connaught Centre for
Biotechnology Research, Molecular Genetics Research, 1755 Steeles
Avenue W., Willowdale, Ontario, M2R 3T4, Canada
LOCATION/Qualifiers
1. .5144
/organism="Haemophilus influenzae"
FEATURES
source


```
Db 3837 GTAAGTGGGTTAAATTAAGGGAATAATTATTTTCGCAGCAGCGAATAAATATGCGAT 3896
Qy 1814 taecgagatgttgattgggattaggggtgogttttgatcaacatcgatttaaatctg 1873
Db 3897 TAGGAAATACATTGAATTTAGGTTTAGGTATTCGGTATCGGTATCTCGTACAAAGCTA 3956
Qy 1874 atgacgtggacacttagccgaacttatcgaaatttggtcttggaaatgggtggattacgc 1933
Db 3957 ATGAATCAACTATTAGTGTGGTAAATTTAAATAATTTCTCTTGGAAATACGTTATGTC 4016
Qy 1934 ttaacaaacagagttgtatcgctttcttatcgcttcaaacgggttttagagtgctg 1993
Db 4017 TAAACCAACGGAATGCTTGATCTTCTTATCGCCTTTCTACTGGATTTAGAAATCCTA 4076
Qy 1994 cattctgaactttatggtaacgtgatcatattgggcttaagaataaagaatatgtgc 2053
Db 4077 GTTTTGTGAAATGTATGTTGGGGTATGTTGGCAATAATAGCGATGTTTATGTAGTA 4136
Qy 2054 aacgcgcgaacgtagccaccagcttagagccagaaaaaatcgactaatcatgagattggag 2113
Db 4137 AA-----TTTAAGCCTGAACACATCTCGTAACCAAGAGTTTGGTC 4175
Qy 2114 ttagtcttaaaaggtcaatttggttaccttgatgtgagctatttccgttaataactataaaa 2173
Db 4176 TCGTCTTAAAGGGGATTTTGGTAAATTTAGATCAGTCATTTTGTAGTAAATGCTATCGAA 4235
Qy 2174 atatga-----ttgcacagcatgtaaaagaataataacaaaaatcacactgtttctata 2227
Db 4236 ATCTTATCGCCTTTGCTGAAGAACTTAGTAAATAATGAACTACTTGAAAGGGCAATATG 4295
Qy 2228 actaccataatcaagatgtacactaaacgggataaaattttagtcgtaaaattgact 2287
Db 4296 GATATCATATATGCACAAATCAAAATAGTTGGCGTAAATATAACTGCGCAATTAGATT 4355
Qy 2288 tacacggtattttatcatgtgcagatggtttttatccatcagttgcttataaacgtg 2347
Db 4356 TTAATGTTTATGGAAACGTTATCCCTACGCTTGGTATGCAACATTTTGTATTAACCGAG 4415
Qy 2348 taaagtaaaagagcgaaactaacccagctcaagactcagatagcgtaaacgactctatc 2407
Db 4416 TAAAGTTAAAGATC---AAAAATCAATGCTGTTTGGCTCGGTAAGCAGATTATTTAT 4472
Qy 2408 tagatggattcagccagcagcatgtgtcttgattcgggtacagctaccaccagaagaaa 2467
Db 4473 TTGATGCCATTTCAGCCCGGCTTATATCATTTGGTTTAGGCTATGATCATCCAAAGTAATA 4532
Qy 2468 aatggggaatttggcattactaccactattctaaagccaaaacgcgcgtaggtggcag 2527
Db 4533 CTGGGGAAATTAATACATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGCTAG 4592
Qy 2528 gcacacgtcatcaggtatcacatcgc-----gttgatttaggtggcaaacctgaccggtt 2580
Db 4593 GACAACGTCATTTGGGTAAACAATTCAAGGAATGTAAATCAACAAGAAAACCTTACTCGGG 4652
Qy 2581 cttagcacaccatgatattaccgg---ttacatcaattataaaaactacacacctagtg 2637
Db 4653 CATGGCATATCTTATAGATGTATCGGGTTATTACATGGCGAATAAAAATATTATGCTCGAT 4712
Qy 2638 gaggaattataatgtgactaatgtaaatattccacttgggaatcagtcgcccgaatccg 2697
Db 4713 TAGGATATATTAATTTATCAACTATCCCTATGTACTTGGGAAGCGGTGGCTCAACAG 4772
Qy 2698 gtgtgaatcgagtaaaccaagacc---ggggtagaatatacactcgatttggcgctccgg 2754
Db 4773 CACAAGGTGCGTCAATCAACATCAAAATGTTGGTAGCTATATACCTCGCTACGCAGCATCAG 4832
Qy 2755 ggagaaattcagtttagcatttgaatgaagtttta 2791
Db 4833 GACGAACACTATACCTTAACATTAGAAATGAATTTCTA 4869
```

RESULT 12

```
I79999
LOCUS 5144 bp DNA PAT 10-JUN-1998
DEFINITION Sequence 105 from Patent US 5708149.
ACCESSION I79999
VERSION I79999.1 GI:3208289
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5144)
AUTHORS Loomore,S., Harkness,R., Schryvers,A., Chong,P., Gray-Owen,S.,
Yang,Y., Mordin,A. and Klein,M.
TITLE Method for producing purified recombinant Haemophilus influenzae
transferrin binding proteins
JOURNAL Patent: US 5708149-A 105 13-JAN-1998;
FEATURES Location/Qualifiers
source 1..5144
BASE COUNT 1813 a 888 c 974 g 1469 t
ORIGIN

Query Match 11.9%; Score 332.6; DB 5; Length 5144;
Best Local Similarity 50.5%; Pred. No. 9.7e-68;
Matches 1353; Conservative 0; Mismatches 1134; Indels 190; Gaps 15;

Qy 173 aaaaagaagtaactgcttggggaagtagtgaaaaattatcaacgaatgagtaaaatc 232
Db 2325 ATAATCAAGTAACGTGACTTGGCAAAATATATAAAACGAGTGAAGATATCAGCCGAGAAC 2384
Qy 233 aaattcttggtatctgatttaactcgtctatgacctggtattctcggtggtggaacaag 292
Db 2385 AAGTATTAATATTCGTGATCTAACAGCTATGATCAGGCATTTAGTGTAGACACAAG 2444
Qy 293 gtgcggtgcaagtagtggtcctgccattgcctaggtgtagataaaaaacccgtgtcagcttac 352
Db 2445 GTCGCGTGCAGTCTCTGGATATTTCTATTCGTGTATGGACAGAAATAGAGTTGCTTTAT 2504
Qy 353 ttgttgatggggtaccacaagcgcagttatcatcagctaggttc----- 398
Db 2505 TAGTAGATGTTTACCTCAAAACCAATCTTATATAGTGCAAAGCCCTTTAGTGTCTCGTT 2564
Qy 399 ----agatgctaattggtgtgcaatgaatgagattgagatgaataaacattcgttcaattg 454
Db 2565 CAGGATATTTGCGCATCTGGTGCATTAATGAATTTGAATGAATAATGTAAGAGCCGCTCG 2624
Qy 455 agttaagcaagagagaagttctgcggaatatggtctggtgcgcagatggtggtgctattg 514
Db 2625 AAATAAGCAAGGGGGGAGTCTCTGAGTATGTTAATGGAGCACACTAGTGTGTTCTGTAA 2684
Qy 515 gtttcgtactaaagatgcgcagagataattataaagaggggcagcattggggcttagata 574
Db 2685 CATTTCAAGCAAAATCCGACGCGATATCTTAGAAGGAGACAAATCATATGGGAATTCAAA 2744
Qy 575 gtaagacctttatgcgcagcaaaaatgccatttttacagttctatccgcgcgctggtg 634
Db 2745 CTAAATATGCTTATTCAGCAAAATAAAGGCTTTTACCCATTCTTTAGCTGTAGCAGGAA 2804
Qy 635 aggcgggtggttttgaagcacttgattgcaactcaccgcacacgcgtaaaagagaccacaaa 694
Db 2805 AACAAAGTGGATTTGAAGGGTGCCTCATTTACACTCAACGAAATTCGGAGAGAAACCCAAAG 2864
Qy 695 ttcattccgaggcaataaaataaacataaatatccgcgtataaacccggtttgaaaaatc 754
Db 2865 TCCATAAAGATGCATTTAAAGGGGTACAAAAGTTATAGCGATTCATCGCC-----ACAACA 2920
Qy 755 gctacgactttacccaaattccgcagagaatgctccctggaggatctctctttaaattg 814
Db 2921 GATAAATCTTCAGGATACTTTGTGATACAAAGGTAGTG-----TCCAATAGGTG 2969
Qy 815 aagatacttgcccaacattagattgtactcctcgtgcaagggttaagtgaacgcgcgata 874
Db 2970 ATGACAAGTGTGCAGC----- 2985
```

Qy 875 atttcccagtgagaaacatttccggaataatacgcctgaaagagcgcgaacacagcttgtagcaga 934
Db 2986 -----CAAACCACTGCAAAAGT 3002
Qy 935 ttccctatcgcagcagcgtctcagcccgaagaataatacccggttaagatgcacattgcac 994
Db 3003 TATCCCCCAAGGAACCGGTAGCGTTTCAGATTATACGGGGCTTAACCGTATCAAAAC 3062
Qy 995 caaaccttttagattacaagagtaattctgttttttagaagtttggctatcaccttcaact 1054
Db 3063 CTAATCCAATGAATATGAAGCCAGCTTGGTTTAAAGAGGAGGTATCATTTTCTG 3122
Qy 1055 cgtctcattatcttggcgcaatcttagaagatacaaaaacgcgtacgatacogtgata 1114
Db 3123 AACCAACTATATTGGTGGTATTTTGAATTCACACACAAAAATTTGATATCGTGATA 3182
Qy 1115 tgcacaacgcagcttactatacaaaagacgatattaacttatacttaggaactatgttt 1174
Db 3183 TGACATTTCCCGCTTATTTAAGATCAACAGAAAAACGGGATGATAGAACTGGCCCTTTT 3242
Qy 1175 atgaagggtgataattatttagatggcttagtgcacccaagagatcccttatgggtgc 1234
Db 3243 ATCCAAACAGATATGGTGCATATCAACGTATTGAGATGGCGGA-----GGCGTTA 3296
Qy 1235 gctatagccatgtgaagttttttagtaacgtcaccaacacgcgtctttagattcacct 1294
Db 3297 ACTATGCAAGTGGGCTTATTTTCGATGAACACCATAGAAAAACAGCGTGTAGGTATTGAAT 3356
Qy 1295 ataaataaacacagagagaataatcgtcgttggtagatataaactc---agtgcgata 1351
Db 3357 ATATTAGCAAAATAAGAACAAAGCGGCATCATTTGACAAAGCAGTGTAAAGTGTCAATC 3416
Qy 1352 aacaagatttgaactatagacggcgtcacatcgtcgtcattgttaggattatcctgtgg 1411
Db 3417 AACAAAAACATCATCTTGACAGTTATATGCGACATACGCAATGGCAGTCTTTATCCTAATC 3476
Qy 1412 tagataaaaattgcgcgcgactttggtaaatcttggctatgtatgcgaactgagcgta 1471
Db 3477 CAAGTAAGAATTGCGCCCGACACTTGATNAACCTTATTCATATNCGTTCGTATAGAA 3536
Qy 1472 ataattaccaagaagaatcgttcattcatttagaattgtataaagcgtataatcgtg 1531
Db 3537 ATGTTTATAAAGAAAAACATAATATGTTGCAATTGAAATTTAGAGAAAAAAATTCACAAA 3596
Qy 1532 gtaaggcgtatttaacaaacccacaaactgaatttagg-----ttgggcttg 1582
Db 3597 ATTGGCTTACTCAACAAATGCTTCAATCTTGTTTGTGATGACTTTACTTTCACGCGCTC 3656
Qy 1583 atcgatttaattcgttatggatcatggggtatgactgcccataatacacaaggcggtt 1642
Db 3657 AGCATAAGATATTATTAACCTGACGTGTACCGCTACGGCAAAATATTATTTCAGGGACAG 3716
Qy 1643 ataccagctacgcggtagagggcgtttagataatccatatttatcgcgcgcgataccac 1702
Db 3717 TTGCTGGTAAACGAGAAATGTTACGAAAAACAACCTTACTTATACFCAAAACCAAAAG 3776
Qy 1703 gcagtattgaacgggtatcttgtgtaataatacacgcgcg-----gacatcttaact 1756
Db 3777 TAGATTTTGTAGGACAAAGATCATTTGTATTAATTATAAGGTAGTCTCTCTAATTTACAGCGACT 3836
Qy 1757 gtgaacgcgtataaataaaggcgtatggcatttggtagcttcgcg---gatcagtga 1813
Db 3837 GTAAAGTGGCTTAAATTAAGGGAAAAATTAATTAATTTGCGAGCAGCAATTAATATGGCAT 3896
Qy 1814 taecgagatgtggtattgggattaggggtgcgtttttagacaacatcgatttaaatctg 1873
Db 3897 TAGGAAATACATTGATTAGTATTAGTATTAGTATTCGTATGAGGTATCTCGTACAAAGCTA 3956
Qy 1874 atgatccgtgacacacttagccgaacttatcgaaatttggcttggaaatgggtgattacgc 1933
Db 3957 ATGAATCAACTATTAGTGTGTTGTAATTTAAAAATTTCTCTTGGAACTAGTGTATTGTCA 4016

Qy 1934 ttaaaccaacagaggttgtgtatcgtctttcttatcgcttattcaaacggttttagagtgctg 1993
Db 4017 TAAACCAACAGGAATGGCTTGATCTTTCTTATCCCTTTCTACTGGATTAGAAATCCTA 4076
Qy 1994 cattctatgaactttatgtaaacgtgatcatatttggccttaaaagataaacgaatatgtgc 2053
Db 4077 GTTTTGTGTAATGTATGGTTGGCGGTATGGTGCAATAATATAGCGATGTTTATATAGGTA 4136
Qy 2054 aacgcgcgcacagctagccaccagcttagagccagaaaaatcgactaaatcatgagattggag 2113
Db 4137 AA-----TTTAAGCCTGAAACATCTCGTAACCAAGAGTTTGGTC 4175
Qy 2114 ttacgtttaaaggctcaatttggttacctttagtgatgtatttccgttaactataaaa 2173
Db 4176 TCGCTCTAAAAAGGGGATTTTGGTAATATTGAGATCAGTCATTTTGTAGTAATGCTTATCGAA 4235
Qy 2174 atatga-----ttgcgacagcatgtaaaagaataatacaaaaaatcacactgtttctata 2227
Db 4236 ATCTTATCGCCTTTTGTGGAAGAACTTAGTAAAAATGGAACCTACTGGAAGGGCAATTATG 4295
Qy 2228 actaccataattcaagatgtagcactaaacgggataaaatttagtcgtctaaatttgaact 2287
Db 4296 GATATCATATATGCAAAATGCAAAATTAGTTGGCGTAATATATTAACTCGCGCAATTAGATT 4355
Qy 2288 tacacgtattttatctatcgtccagatggtttttttattcatcagttgttctataaacgtg 2347
Db 4356 TTAATGTTTATGGAACAGTATTCCTTACGTTGGTATGCAACATTTGCTTATTAACCGAG 4415
Qy 2348 taaaagtaaaaggcggaactaaacgcgactcaagactcgtatagcgtataacgactcctattc 2407
Db 4416 TAAAGTTAAAGATC---AAAAATCAATGCTGTTTGGCTCCGTAAAGCAGTTATTAT 4472
Qy 2408 tagatggattcagcgcagcgtatgttggattcgcgtacgataccaccagaagaaa 2467
Db 4473 TTGATGCCATTACCCCGCGCTTATATCATTTGTTTAGGCTATGATCATCCAAAGTAATA 4532
Qy 2468 aatgggaattggcattactaccacctattctaaagccaaaaacgcgatgaggtggcag 2527
Db 4533 CTGCGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATGTAG 4592
Qy 2528 gcacacgtcatcgcggtatcatcgcg-----gttatttaggtggcaaacgtgaccggtt 2580
Db 4593 GACAACGTGCTATTGGTAAACAATTAAGGAATGTAAATCAACAAGAAACTTACTCGGG 4652
Qy 2581 ctgtgtacacccatgatattacgcg---ttacatcaattataaaaaatacaccttagtg 2637
Db 4653 CATGGCATATCTTAGATGATTCGGGTTATTACATGCGGAATAAAAAATATTATGCTTCAT 4712
Qy 2638 gaggaatttaataatgtacaaatcgttaattccacttgggaatcagtcgccaatccg 2697
Db 4713 TAGGATATATTAATTTATTCAACTATCGCTATGTTACTTGGGAAGCGGTGGTCAACAG 4772
Qy 2698 gtgtgaatgcagtaaaccaagacc---ggggtagcaattacactcgtatttggcctcgcg 2754
Db 4773 CACAAGGTGCGGTCAATCAACATCAAAATGTGGTAGCTATATCTCGTACGACGATCAG 4832
Qy 2755 ggagaaatttcagtttagcatttgaagaatttga 2791
Db 4833 GACGAACCTATACCTTAACATTAGAAATGAATTTCTA 4869

RESULT 13

HIU10882
LOCUS 4927 bp DNA BCT 04-NOV-1995
DEFINITION Haemophilus influenzae transferrin binding protein 2 (tbpB) gene
and transferrin binding protein 1 (tbpA) gene, complete cds.
ACCESSION U10882
VERSION U10882.1 GI:747921
KEYWORDS Haemophilus influenzae.
SOURCE Haemophilus influenzae.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.


```
Db 3081 TGACATTCCGCGTTATTTAAGGCCAACAGAGGATTTACAAAGTGGCCCTTTT 3140
QY 1175 atgaagggaataatatttagatgcttagtggtcaagccaagatcccttatgggtgc 1234
Db 3141 ATCCAAAGCAGAT- ----TATGGTCATATCAACATATTTGGTGATGGCAGAGCGCTTA 3194
QY 1235 gctatagccatgtgaagttttttgatgaacgtaccacaacgcgtgttaggattcacct 1294
Db 3195 AATATGCAAGTGGCTTATTTTCGATGAACACCATAGAAAACACGCGTGTAGGTATTGAAT 3254
QY 1295 ataaataaaccagagaataa ---tcgctgggttgatagcattaaactcagtcgagata 1351
Db 3254 ATATTTACGAAATTAAGAACAAAGCGGCATCATGTGCAAGCGGTGTAAAGTCTAATC 3314
QY 1352 aacaagatttgaactatatagcgcgtacatcgcttgcatgttagcgatttacctgtg 1411
Db 3315 AACAAACATCATACTTGCAGATTTATGCGACATACGCATTCGAGCTTTATCCTAATC 3374
QY 1412 tagataaaattccgcgccactttggataaatttggctatgtatcgaactgagcgta 1471
Db 3375 CAAGTAAGAATTGCGGCCAACACTTGATAAACCTTTATCATACTATCATCTGTATAGAA 3434
QY 1472 ataattaccagaagcatcgctcattcatttagaattttagaattgataaagcgtcaaatgctg 1531
Db 3435 ATGTTTATAAGAAAAACATAACAGGTGCAATTTGAATTTAGAGAAAAAATTCAACAAA 3494
QY 1532 gtcaaggcgtatttaacccaaacccacaaactgaatttaggttggttgcttgatcgatta 1591
Db 3495 ATTGGCTTACTCATCAAAATGTCCTTCAATCTTGTTTGA ---TGACTTTACTTCCGCAC 3551
QY 1592 attogcttatggatcatg9gggatagactgcoccaataataccaaggcggttataccagct 1651
Db 3552 TTCAGCATAAAGATTATTAACTCGAGCTGTATCGCTACGGCAAGTAGTATTTCAGAGA 3611
QY 1652 acgcggtagaggcggttagat-----aatccatatatttatcgccgcgac 1699
Db 3612 AACGTGTGAAGCAAGAAAGATGTTTACAATCAAGCTCTTACTTATACCCACACCAA 3671
QY 1700 cagcgatttgaacggtatcttctgttaataacacg-----cggcgacatcttaa 1753
Db 3672 AAGCAGAGTTGGTAGGAGGAGACTTTGTAAATATCAAGTAGTACTCTTAATACAGTG 3731
QY 1754 actgtgaaccgcgtaaattaaaggcgatagccattttgttagcttcgc---gatctag 1810
Db 3732 ACTGTAAAGTCGGTTAATTAAGGGAAAAATTATTATTCGACGACGCAATAATATGG 3791
QY 1811 tgataagcgatgtggtatttggatttgggtgctgcttttgatcaacatcgatttaaat 1870
Db 3792 CATTAGGGAATACGTTGATTTAGGTTTAGGTATGAGGTATGACGTATCTCGTACAAAAG 3851
QY 1871 ctgatgctcgtggacacttagccgaacttatgaaattggtcttgaatggtgggatta 1930
Db 3852 CTAAATGAATCAACTATTAGTGTGGTAAATTTAAATAATTTCTCTGGAATACTGGTATG 3911
QY 1931 cgcttaaaccaacagagtttggatcgctttcttcattcgcatttcaaacggttttagatgc 1990
Db 3912 TCATAAACCAACAGGAATGCTTGATCTTCTTATCGCTTCTTCTACGTGATTAGAAATC 3971
QY 1991 ctgcattctatgaactttatggttaaacgctgatcataatgggctttaaagaatacgtatg 2050
Db 3972 CTAGTTTTGTGAAATGATGTTGGGTTGGGTATGGTGAAGGATACCGATGTTTA----- 4026
QY 2051 tgcaacgcgcgaacgtagccaccagttagaccagagaaaaatcgactaatcatgagattg 2110
Db 4027 -----TATAGTAAATTTAAGCCTGAAGCATCTCGTAACCAAGAGTTTG 4070
QY 2111 gagtagctttaaggttcaatttgggtaccttgatgagctatttctcgtaataaactata 2170
Db 4071 GTCTCGCTCTAAAGGGGATTTGGTAAATATTGAGATCACTCATTTTGAATAGCTTATC 4130
QY 2171 aaatatga-----ttgcgacagcatgtaaaagaataatacaaaaaatcacactgttct 2224
```

```
Db 4131 GAAATCTTATCGCTTTGCTGAAGAACTTAGTAAAAATGGAAGTACTACTGGAAGGCAATT 4190
QY 2225 ataactaccataatattcaagatgtagcactaaacgggataaatttagtcgctaaatttg 2284
Db 4191 ATGGATATCATATATGACAAAATGCAAAATAGTTGGCGTAAATATATACCTGCGCAATTAG 4250
QY 2285 acttacacggtattttatctatctgctccagatggttttttattcatcagttgctttataacc 2344
Db 4251 ATTTTAATGTTTATGAAACGATATCCCTACGTTGGTATGCAACATTTTGTCTTATAACC 4310
QY 2345 gtgtaaaagttaaaagagcggaactaaacgactcaagactcgatagcgtaaacatccta 2404
Db 4311 GAGTAAAGTTAAAGATC-----AAAAATCAATGCTGTTTAGCTTCGCTAAGCAGTTATT 4367
QY 2405 ttctagatcgattcagcgagcagcgtatgcttgcattgcgttcgctacgatacccccagaag 2464
Db 4368 TATTTGATGCCATTCAGCCAGCGGTATATCATTTGTTTAGGCTATGATCATCCAAAGTA 4427
QY 2465 aaaaatggggaattggcattactaccactatttctaaagccaaacgcgcgtagggtg 2524
Db 4428 ATACTTGGGAATTAAGACAAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGC 4487
QY 2525 caggcacacgtcatcagcgatacatgc-----gttgatttaggtggcaaacgtgacg 2577
Db 4488 TAGGAAAACGTGATTTGGGTAAACAATTCAGGAATGTAAAATCAACAAGAAAACTTTACTC 4547
QY 2578 gtcttggtcacccatgatattaccgg---ttacatcaattataaaacacacacctac 2634
Db 4548 GGGCATGCCATNTCTAGATGTTATCGGTTTATTCATGTTGGAATGAAATGATTTTGTCTCC 4607
QY 2635 gtggaggaatttaataatgtgactaaatcgtaaatattccacttgggaatcagtcgccaat 2694
Db 4608 GATTAGGAGTATATAATTTTAAACTATCGTATGCTATGCACTTGGGAAGCGTGGCTCAA 4667
QY 2695 cgggtgtaatgtagtaaaccaagacc---gggtagcaattacactcgtattggcgctc 2751
Db 4668 CAGCACAAAGTCGGTCAATCAACATCAAAATGTTGTAATCTATCTACTCGCTACGAGCAT 4727
QY 2752 cggggagaatttcagtttagcatttgaatgaatttta 2791
Db 4728 CAGGACGAAACTATACCTTAACATTAGAAATGAAATTTCTA 4767
```

RESULT 14

```
HIU15052
LOCUS HIU15052 5009 bp DNA BCT 02-APR-1996
DEFINITION Haemophilus influenzae Minna transferrin binding protein 1 (tbp1)
and transferrin binding protein 2 (tbp2) genes, complete cds.
ACCESSION U15052
VERSION U15052.1 GI:1223940
KEYWORDS
SOURCE Haemophilus influenzae.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE 1 (bases 1 to 5009)
AUTHORS Loosmore, S.M., Yang, Y.P., Coleman, D.C., Shortreed, J.M.,
England, D.M., Harkness, R.E., Chong, P.S. and Klein, M.H.
TITLE Cloning and expression of the Haemophilus influenzae transferrin
receptor genes
JOURNAL Mol. Microbiol. 19 (3), 575-586 (1996)
MEDLINE 96228706
REFERENCE 2 (bases 1 to 5009)
AUTHORS Loosmore, S.M.
DIRECT SUBMISSION
TITLE Submitted (23-SEP-1994) Sheena M. Loosmore, Connaught Centre for
Biotechnology Research, Molecular Genetics Research, 1755 Steeles
Avenue W., Willowdale, Ontario, M2R 3T4, Canada
FEATURES
Location/Qualifiers
1..5009
/organism="Haemophilus influenzae"
/strain="Minna"
/specific_host="Homo sapiens"
```

```

/db_xref="taxon:727"
/clone="DS-712-1-3"
121..2103
/gene="tbp2"
121..2103
/gene="tbp2"
/codon_start=1
/transl_table=11
/product="transferrin binding protein 2"
/protein_id="AAC43926.1"
/db_xref="GI:1223941"
/translation="MKSVPLISGLSLFLLSACSGGSDVDVNSPTSSKPRYQDDTS
NQRKSNLKLPIPSLGGGKMLVAQLNRKNKPSFLNEDDIYSFSLSTIEKDVKN
NKGADLIGSIDEPSTNPPEKHGOKYVYSGLYYTPSLNDKSNKFLVGYGVAFY
YGNKATNPLPVGVAAYKGTWDFITATKNGKRYPLLSNGSHAYIRSDALPEIDLEND
KNGDGLISERSADRGTKRLTGQLSYTRKNNPIERKKLIDIDIASIYISNRFQTV
KPTKDSSEHPSTSEGLEGGFYGNAELGGKFLATDNRTVTFVSFAKETETKEAL
SKETLDGLIYFTSTKTDANTSTAANTTTDTANTITDEKRFETIDESSGEA
DYLLIDKYPILPDKNTDFISSKHHTVGNKRYKVEACCSNLSYVKGFWYEDPLKE
KETETETEKDEKEKDKOKETAATNTYYOFLGLHRTPKDDIDPKTGSAYHG
SNFGYITDKTYSVSPSGDKKROKNAVAEENFPAEKKLFGELKRHDTPNPVSIENF
NNSNAPTGTATATNFIIDGKNSQNTFINITIKVNGAFYGPKASELGGYFTYNGNS
TATNSSTSVSSSSNKNARAIVFEGARQVETTK"
2117..4855
/gene="tbp1"
2117..4855
/gene="tbp1"
/codon_start=1
/transl_table=11
/product="transferrin binding protein 1"
/protein_id="AAC43927.1"
/db_xref="GI:1223942"
/translation="MTKKPYFRLIISCLLIICYVKAETQSLIKDTKEALISSEVDQST
ESELSTISYAEKINDRDNEDVIGLGIKIKTSESISREQVLNIRDLTRDGRGISVVE
QGRGSSGYSIRGMNRNVALLDVGLPQTQSYVQSPLVARSYGISGTGAINELIENV
KAVEISKGSSSEIYAGLANSYTFQSKAADILEGDKSWGITKNAYSKNKGFTHS
LAVAGQGGFEGALYQALSIETQVHKDAALGVQSYDLRIATTDKSSGFYVIOGBCP
NGDKCAAKPAPLTSQTSVSYDVTGANRIKPNPKMYESOSWFURGVHSEQHYI
GGIFETQOKFDLDMTFAYLSPTERRDDSSRFYPMODHGAYOHIEDRGVKNYASG
LYFDEHHRQRQVGEIYIENKNKAGIIDKXVLSANQNIILIOSYMRHTSCSLYPNPSK
NCRPTLDKPYISYRDNRYIEKKNMLQNLERKKIQOONMLTHQIVFNGLDFDFTSALQ
HKDYLTVIATADSIPRKPGETGKPRNGLSQPLYLPKPEYFAQGDHCFTQSSSN
YRDKCVLLIKKNYFAARNMALGYVDLGIIRYDVSRTKANESTISGVFKNFSW
NTGIVLKPTWELDSLRLSTGPNPSFSEMYGWRVGGKNDVYVGVFKPETSBNQFEG
LALKGDFGNEIISHFNAYENLFAFAELSKNGTGKNGYHNAQNAKLVGNITLAQL
DFNLWKRPYGHYATFAYNOVKVQDKINAGLASVSSILFDIAQPSRTIIGLGDHP
SNWGINTFMTQSKNSQNELKRALGNNSRDVSKRLKRAWHILDIYSVGMANKN
IMLRGLIYNLFNRYVTWEAVRQTAGAVNQHVQNGSYTRYASGRNRYLTLEMKP"
BASE COUNT 1809 a 836 c 934 g 1430 t
ORIGIN

Query Match 11.9%; Score 331.2; DB 2; Length 5009;
Best Local Similarity 50.5%; Pred. No. 2.1e-67;
Matches 1354; Conservative 0; Mismatches 1133; Indels 193; Gaps 15;

QY 173 aaacaggaataactggttggggaagaatagtgaaaaattatcacgaaatagtaaaaaac 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2307 ATATGAAGTAACATGGACTTGGCAAAATATCAAAAACCTAGTGAAGATATCAGCCGAGAAC 2366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 233 aaatctgtgattcgattgaactgcctatgacctggtatttcggtggtggaacaag 292
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2367 AAGTATTAATATTCGTATACACGCTATGATCCAGGGATTTTCAGTTGTAGAACAAAG 2426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 293 gtccggtgcaagtagtggtatgccattcgaggtgtagataaaaaaccgtgtcagcttac 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2427 GTCGCGTGCAAGTCTCGGATATTCATTCGTGTATGGACAGAAATAGAGTTGCTTTAT 2486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 353 ttgttgatgggtaccacaagcgacagttatcatcacgtaggttc----- 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2487 TAGTAGATGGTTTACCTCAACGCAATCTTATGATGAGCAAGCCCTTATGTCGTGTT 2546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 399 ----agatgctaaagggtgcaattaatgagattgagattgagaaacattcgttcaattg 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 3519 ATGTTTATAAGAAAAACATAATATGTTGCAATTGCAATTTAGAGAAAAAATTCACAAA 3578
Qy 1519 --gocataatgctgtgcaagcgctatttaacaaacccaaacgaactgaatttaggttgg 1576
Db 3579 ATTGGCTTTACTCATCAAAATGTCTTCAATCTGTGTTTGGTATGACATTTACTTCAGGCTTC 3638
Qy 1577 gctttgatcgatttaattcgcgtattggtatcgatcgatgggataatgactgcccaataaccagaag 1636
Db 3639 AGCATAAAGATATTATTAACCTGCACGTGTTATCGCTACGGCAGATAGTATTCAGGAAAC 3698
Qy 1637 gcgggttaccagctaccgcggtgagggcggttttagataaattccatattatttgcgcgcg 1696
Db 3699 CTGGTGAACCTGGTAAACCAAGAAATGTTGCCATCAACACCTTACTTATACCCAAAC 3758
Qy 1697 atccacgcagttattgaacggttatctttgtgtaataatacacgcggc-----gacatct 1750
Db 3759 CAGAGCCATATTTTCAGGACAAGATCATTTGTAATTTATCAAGGTAGCTCCTCTAATTACA 3818
Qy 1751 taaactgtgaaccggtataaataaagcgatagccattttgttagcttcgcg---gac 1807
Db 3819 GAGACTGTAAAGTGGGTTAATTAAGGGAATAATTTATTTTCGCACGACGCAATAATA 3878
Qy 1808 tagtgataagcgagtgatgtgatttggttaggggtgcttttgatcaacatcgattta 1867
Db 3879 TGGCATTAGGGAATACGTTGATTTAGGTTTAGGTATTTCGGTATGACGTATCTCGTACAA 3938
Qy 1868 aatctgatgaccgctggacacttagccgaacttatcgaaatttgcttgggaatggtggga 1927
Db 3939 AAGCTAATGAATCAACTATTAGTGTGTGTAATTAATTTAAAAATTTCTCTGGAACTACTGCTA 3998
Qy 1928 ttacgcttaaaccaacagagtttgatcgctttcttcttcttcgatttcgaacaggttttagag 1987
Db 3999 TTGTCATAAAAACCAACGAATGGCTTGATCTTCTTATCGCTTTCTACTGGATTAGAA 4058
Qy 1988 tgcctgcatctctgaactttatgtaacgctgatacgtatggtggctttaaataacgaat 2047
Db 4059 ATCCTAGTTTTCTGAATGTATGTTGGCGGTATGGTGGCAAGAAATGACGAGGTTTATG 4118
Qy 2048 atgtgaacgcgcgaacgtagccaccagtttagagccagaaaaatcgactaatcatgaga 2107
Db 4119 TAGGTAAA-----TTTAAGCCGTAAACATCTCGTAACCAAGAGT 4157
Qy 2108 ttgagttagctttaaaggtcaatttggttacccttgatgtagctatttcggttaataaact 2167
Db 4158 TTGGTCTCGCTCTAAAGGGGATTTTGGTAATATTAGATCAGTCATGTCATTTAGTATGCTT 4217
Qy 2168 ataaaaataatgtgcagacgatgaaa---agaataatacaaaaaatacacagtcttct 2224
Db 4218 ATCGAAATCTTATCGCCCTTTCGTGAAGAACTTAGTAAAAAATGGAACTGGAAGGGCAAT 4277
Qy 2225 ataaactacaataattcaagatgtagcactaaacgggataaaatttagctgctaataattg 2284
Db 4278 ATGGATATCATATGCACAAAATGCAAAATAGTTTGGCGTAAATATATACTGCACANTTAG 4337
Qy 2285 acttacacggttatttatactatgctgcagatggtttttattcatcagttgcttataacc 2344
Db 4338 ATTTTAAATGGTTTATGGAAAGCTATTCCCTACGGTGTGTATGCAACATTTGCTTATAACC 4397
Qy 2345 gtgtaaaaagtaaaagcggaactaaacgcgactcaagctcgatagcgtaaacgatccta 2404
Db 4398 AAGTAAAAAGTTAAAGATC---AAAAATCAATGCTGGTTTAGCCCTCCAGGAGTTATT 4454
Qy 2405 ttctagatgcattcagcagcacgctatgtgcttggtatcggttcagcatcacccagaag 2464
Db 4455 TATTGTATGCCATTCAGCCACGCCCTTATATCATTTGGTTTAGGCTATGATCATCCAGTA 4514
Qy 2465 aaaaatggggatggcattactaccactattctaaagcgaaccccgatgaggtgg 2524
Db 4515 ATACTTGGGAATTAATACAATGTTTACTCAATCAAAAAGCAAAATCTCAAAATGAATTGC 4574
Qy 2525 caggcacagtcatacagatacatcgc-----gttgatttaggtggcaactgaacg 2577
Db 4575 TAGGAAAACGTCATTTAGGTAAACAATTCAGGGGATGTAATAATCAACAAGAAAACCTTACTC 4634

Qy 2578 gtctttgtacacccatgatattaccgg---ttacatcaattataaaaaactacaccttac 2634
Db 4635 GGGATGSCATATCTTAGATGTATCGGTTATTACATGCGAATAAAAAATATTATGCTTC 4694
Qy 2635 gtgaggaatttataatgtgactaatgtaaatattccacttgggaatcagtggcceaat 2694
Db 4695 GATTAGGGATATAATATTTATTCACATATCGCTATGTTTGGGAACGGGTGCGTCAA 4754
Qy 2695 ccggtgtgaatgcagtaaaaccaagacgg---gggtagaattacactcgatttggcgctc 2751
Db 4755 CAGCACAGGTGCGGTCAATCAACATCAAAATGTTGGTAGCTATATCTGCTACCGAGCAT 4814
Qy 2752 cggggagaaatttcagtttagcatttgaattgaattgaatttta 2791
Db 4815 CAGGACGAACATATACCTTAACTTAACATTAGAAATGAATTTCTA 4854

RESULT 15

LOCUS I79989 5009 bp DNA PAT 10-JUN-1998
DEFINITION Sequence 3 from patent US 5708149.
ACCESSION I79989
VERSION I79989.1 GI:3208279
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5009)
AUTHORS Loosmore,S., Harkness,R., Schryvers,A., Chong,P., Gray-Owen,S.,
Yang,Y., Mordin,A. and Klein,M.
TITLE Method for producing purified recombinant Haemophilus Influenzae
transferrin binding proteins
JOURNAL Patent: US 5708149-A 3 13-JAN-1998;
FEATURES Location/Qualifiers
source 1..5009
BASE COUNT 1809 a 836 c 934 g 1430 t
ORIGIN

Query Match 11.98; Score 331.2; DB 5; Length 5009;
Best Local Similarity 50.5%; Pred. No. 2.1e-67;
Matches 1354; Conservative 0; Mismatches 1133; Indels 193; Gaps 15;

Qy 173 aaaaagaagtaactgcttgggaaagtagtaaaaattatcacgaatgataaaatc 232
Db 2307 ATAATGAAGTAACCTGGACTTGGCAAAATTTACAAACTAGTGAAGTATCAGCCGAGAAC 2366
Qy 233 aaattcttggtattcgtgatttaactcgctatgacctggttatttcggtggtgaacaag 292
Db 2367 AAGTATTAAATATTCGTGATCTAACACGCTATGATCAGGGATTTCACTTGTAGAACAAG 2426
Qy 293 gtccggtgcgaagtagtggtctaccattccaggtgaggtgtagataaaaaacggtgcagcttac 352
Db 2427 GTCGGGTGCAAGTCTGGATATTCTATTCGTGTTATGATGACAGAAATAGAGTTGCTTTAT 2486
Qy 353 ttgttgatgggtaccacaagcgacagtatcatcagctaggttc----- 398
Db 2487 TAGTAGATGGTTTACCTCAAAACGCAATCTTATGTAGTCAAAAGCCCTTTAGTTGCTCCTT 2546
Qy 399 ---agatgctaaatgggtgctaaatgattgagattgagatgaacaacattgctcaattg 454
Db 2547 CAGGATATCTCGCACCTGGTCAATTAATGAATTAATGAATGAATAATGAAGCCGCTCG 2606
Qy 455 agttaagcaagggaggaagttctcgcgaatatggtctcgtgcgcatggtggtgctattg 514
Db 2607 AAATAAGCAAGGGGGGAGTCTTCTGAGTATGTTAATGGAGCAGCTAGCTGTTCTGTAA 2666
Qy 515 gtttcgtactaaagatgcgcagatattttaaagagggcgacgattgggcttagata 574
Db 2667 CATTTCAAGCAAAATCAGCAGCCGATATCTTAGAAGGAGAGACAAATCATATGGGGAAATTCAAA 2726

Qy 575 gtaagaccttattgacagcaaaatagccattttttacagctctatgcgacggctggtg 634
Db 2727 CTAAATGCTTATTCAAGCAAAATAAAGGCTTTACCATTCTTTAGCTGACAGAA 2786
Qy 635 aggcgggtggttttgaagcaactgtttattgcaactaccgcacacggttaagagacaaaa 694
Db 2787 AACAAAGTGGATTTGAAGGGTAGCCATTACACTCAACGAATTCATTTGAACCCCAAG 2846
Qy 695 ttcatccagagcaaaataaataaataatttcggtatatttcggtatataacggctttgaaatc 754
Db 2847 TCCATAAAGATGCAATTAAGAGCGGTACAAAGTTATGATCGATTATCGCC-----ACAA 2902
Qy 755 gctacgactttaccacaattccgcacagaatgctcctgagagatcctctttaaattggtg 814
Db 2903 GATAAATCTTCAGGATACTTTGTGATACAAAGTGAGTG-----TCCAAATGGTG 2951
Qy 815 aagatacttgcaccaatattagattgtactcctcgtgcgaagggttaagttgacgcgata 874
Db 2952 ATGACAAGTGTGCAGC----- 2967
Qy 875 atttcccgagtgaacaatttcgggaataatacgcctgaagagcgcaaacagattgagcaga 934
Db 2968 -----CAAGCCACTGCGACTT 2984
Qy 935 ttccttattcgactgagcagctctcagcccagaataataccggttaagatcgcatgac 994
Db 2985 TATCACCCANAGCGAACCCTGTAAGCGTTTCAGATTATACGGGGCTTAACCGTATCAAC 3044
Qy 995 caaaccttttagattacaagagtaattcgtttttttatgaagtgttggtatcaactcaact 1054
Db 3045 CTAATCCAATGAATATGAAGCCAGTCTTGTTTAAAGAGGAGGTATCATTTTCTG 3104
Qy 1055 cgtctattattcgtgcgaactcttagaagatacaaaaaacgcgtacgatacgtgata 1114
Db 3105 AACAACTATTATTGGTGGTATTTTGAATTCACACACACAAATTTGATATCCGTGATA 3164
Qy 1115 tgaaccccgacttactatacaaaagcagatattcaacttcaacttaggaactggtt 1174
Db 3165 TGACATTTCCCGCTATTATTAGCCCAACAGAAAGCGGGATGATAGTAGTCGTCTTTT 3224
Qy 1175 atgaagggtataatttttagatggcttagtgttcaagccaagatcccttatgggtgc 1234
Db 3225 ATCCAATGCAAGATCATGGTCATATCAATATATTGAGGATGGCA-----GAGGCGTTA 3278
Qy 1235 gctatagcattggaagttttttgatgaagtcaccacaacacgctgtttaggattcaact 1294
Db 3279 AATATGCAAGTGGGCTTTATTTCGATGAACACCATAGAAAAACGCGTGTAGGTATTGAAT 3338
Qy 1295 ataaataaaacagagataatcgtggttggatagcattaaactc---agtcgggata 1351
Db 3339 ATATTTACGAAATAAAGCAAAAGGGGCGCATCAITGACAAGACGAGTGTAAAGTCTAATC 3398
Qy 1352 aacaagatttgaactatataagcgggtcacatcgttcattgttagcattatcctgtg 1411
Db 3399 AACAAACATCATACTTGACAGTTATATGCGACATACGCAATTGCGAGTCTTTATCCTAATC 3458
Qy 1412 tagataaaaattgcgcgcgactttgggataaaacttgggtctctatgatacgaactgagcgt 1471
Db 3459 CAAGTAAGAAATGCGCCCAACACATGATAAACCTTATTTCATCTATCTCTGTATAGAA 3518
Qy 1472 ataatataaagaagaacatcgtgctcattcatttagaatttgataaa----- 1518
Db 3519 ATGTTTAAAGAAAAACATATATTGTCGAATTTGAATTTAGAGAAAAAATTCNACAAA 3578
Qy 1519 --gctcaaatgctggtgaagcgctatttaaccacaacccacaactgaattagggtg 1576
Db 3579 ATTGGCTTACTCATCAAAATGCTTCAATCTGTTGTTTGTGACATTTACTTCAGGCGTTC 3638
Qy 1577 gctttgacgatttaattcgtcttattgatacgtggtggtatgactgcccataataccaaag 1636
Db 3639 AGCATAAAGATTATTAACTCGACGTGTTATCGCTACGGCAGATAGTATTCAGGAAAC 3698
Qy 1637 ggggttaaccagctaccgcggttagaggcggttttagataaattacactcgtatttgcgcgc 1696

Db 3699 CTGCTGAACCTGCTAAACCAAGAAATGGTTGCAATCACAACCTTACTTATACCCAAAAC 3758
Qy 1697 atccacgagatttgaacacggtatcttctgtataataatacacgcgc-----gacatct 1750
Db 3759 CAGAGCCATATTTTGCAGGACAAGATCATTTAATTAAGGTAGCTCTCTAATTACA 3818
Qy 1751 taaactgtaacgcgtataaataaagcagcagcattttgttagcttcgc---gactc 1807
Db 3819 GAGACTGTAAGTCGGGTTAATAAGGGAATAATTTATTTCCGACGACGCAATTAATA 3878
Qy 1808 tagtgaalaagcgagtagtggattggggttaggggtgcgttttggatcaacatcgattta 1867
Db 3879 TGGCATTAGGGAATACTGTTGATTAGTATTAGTATTCGGTATGACGTATCTCGTACAA 3938
Qy 1868 aatcgtatgacgcgtgagacacttagccgaacttatcgaaatttgccttgaattggtgga 1927
Db 3939 AAGCTAATAATCAACTATTAGTGTGGTAAATTTAAAAATTTCTCTTGGAAATCTGTA 3998
Qy 1928 ttacgcctaaacacagagttgtatcgcttcttctatcgacttcaaacggttttagag 1987
Db 3999 TTGTCATAAACCAACCAAGGATGGCTTGATCTTTCTTATCGCTTCTACTGGATTTAGAA 4058
Qy 1988 tgcctgacttctatgaacttttagtaacgctgacatattgggctttaaagataaacgaat 2047
Db 4059 ATCCTAGTTTCTGAAATGATGTTGGCGGTATGTTGGCAAGAAATGACGAGGTTTATG 4118
Qy 2048 atgtgaacgcgcgcacgtagccaccagtttagagcagacagaaaaactgactaatacataga 2107
Db 4119 TAGTAA-----TTTAAAGCCCTGAACATCTCGTAACCAAGAGT 4157
Qy 2108 ttggagttagctttaaaggtcaatttggttaccttgatgagctatttccgtataaact 2167
Db 4158 TTGCTCTCGCTCTAAAAGGGGATTTTGGTAATATTGAGATCAGTCATTTTAGTAATGCTT 4217
Qy 2168 ataaaatatgattgcacagcatgtaaa---agaaataacaaaaaacacactgtttct 2224
Db 4218 ATCGAATCTTATCGCTTTGCTGAAGACTTAGTAAATAATGGAACCTGGAAGGCAATT 4277
Qy 2225 ataactaccataattcaagatgtagcactaaacgggataaaatttagtgcgttaaaattg 2284
Db 4278 ATGGATATCAATAATGCAAAAAATAGTGTGGCTAAATATAAATACTGCACAAATAG 4337
Qy 2285 acttaacggtattttatctatcgtccagatggttttttatttcatcagttgcttataacc 2344
Db 4338 ATTTAATGGTTTATGAAACGTTATCCCTACGGTTGGTATGCAACATTTCTTATATACC 4397
Qy 2345 gtgtaaaagtaaaagagcggaactaaacccgactcaagactcgatagcgtaaacgactccta 2404
Db 4398 AAGTAAAGTTAAAGATC---AAAAATCAATGCTGGTTTAGCCTCGTAAAGAGTTATT 4454
Qy 2405 ttctagatgcgattcagccagcagctatgctgtgattgcgtgattgcgtacgataccacagaag 2464
Db 4455 TATTGTATGCCATTACAGCCAGCGGTTATATCATTTGTTTGGCTATGATCATCCAACTA 4514
Qy 2465 aaaaatgggaattggcattactaccacttctaaagcacaacacgagcgtatagggtg 2524
Db 4515 ATACTGGGGAAATTAATCAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGC 4574
Qy 2525 caggcaacgctcatcaacggatatactgcg-----gttgatttaggtggcaactgacgcg 2577
Db 4575 TAGGAAAACGTGATAGGTAAACAATTAAGGGATGTAATAACAAAGAAAACTTACTC 4634
Qy 2578 gttcttggtaacccatgatatattaccg---ttatcataaattataaaactacacacttac 2634
Db 4635 GGGATGCGCATCTTAGATGATCGGTTATTACATGGCGGAATAAAAAATTTATGCTTC 4694
Qy 2635 gtgaggaattataatgtgactaatgtaaatattccacttgggaatcagtcgccaat 2694
Db 4695 GATTAGGATATATATTTATTCACTATCGCTATGTTACTTGGGAACGGGTGGTCAAA 4754
Qy 2695 ccggtgtgaatgcagtaaaacacagacgcg---gggtagaattatcactcgtatttgcgcgc 2751

Db 4755 CAGCACAGGTGGGTCAATCAACATCAAAATGTTGGTAGCTATACTCGCTACGCACCAT 4814

Qy 2752 cgggggaaatttcagtttagcatttgaagtgaagtta 2791

Db 4815 CAGGACGAACTATACCTTAACATTAGAAATGAATTCTA 4854

Search completed: September 14, 2000, 00:15:09
Job time: 3752 sec

THIS PAGE BLANK (USPTO)

Query Match		99.18;	Score 2766;	DB 1;	Length 2784;		
Best Local Similarity		99.78;	Pred. No. 0;	Mismatches	0;	Indels	8; Gaps
Matches 2784;		Conservative	0;				1;
Qy	1	atgataatgaatatcatcatttcctgcgtatccactgttgctcttaaacagtgattatttgct	60				
Db	1	ATGATAATGAATATATCATCATTTTCGCTATTTCACCTGTTCCTTTAACAGTGTATTTGCT	60				
Qy	61	ctttctcatcactcagtgctgcgactgaaataaaaaaaacgaagaaaaataacgatctta	120				
Db	61	CTTTCTCATTCATACGGTGCCTGCGACTGAAATAAAAAATCGAAGAAAATAACGATCTA	120				
Qy	121	gctgtctcagatgaagttattgtacagagagccattatgctcagcagctcacaacgaa	180				
Db	121	GCTGTCCTGAGATGAAGTTATTGTGACAGAGAGCCATTATGCTCACGAACGTCGCAAAACGAA	180				
Qy	181	gtaactggcttgggaaagtagtgaaaaattatcacgaaatagtgtaaaaaatcaaatcttt	240				
Db	181	GTAACCTGGCTTGGGAAAGTAGTGAATAATATCACGAAATGAGTAAAAATCAAAATCTTT	240				
Qy	241	ggatttcggtatttaactcgcgtatgaccttggtatttctgggtggaaacaggtcgcggt	300				
Db	241	GGTATTCGTGATTAACTCGCTATGACCTGTGTTTTCGGTGGTGAACAAGGTGCGGGT	300				
Qy	301	gcaagtagtgctatgccattcgaggtgtagataaaaaaccgtgtcagcttacttcttgat	360				
Db	301	GCAGTAGTGCTATGCCATTTCGAGGTGTAGATAAAAACCGTGTACAGCTTACTTTGTGAT	360				
Qy	361	gggtaccacaagcgacagttatcatacgcgtaggttcagatgctaatgggtgggtgcaatt	420				
Db	361	GGGTACCACAAGCGCACAGTTATCATACGCTAGGTTCAGATGCTAATGGTGGTGCAAAT	420				
Qy	421	aatgagatgagtgataaaaaattcgttcaattagtttaagttaacaaaggagcaagttctgcg	480				
Db	421	AATGAGATTGAGTATGAATAACATTGTTCAATTGAGTTAAGCAAAAGGAGCAAGTCTTCGC	480				
Qy	481	gaatatgctctggcgcatgggtgctgtattggtttcttactaaagatgcgcagat	540				
Db	481	GAATATGCTCTGGTGCCGANTGGTGCTATNTGGTTTTCTACTTAAAGATGCCGAGAT	540				
Qy	541	attattaagaggggcagcattggggcttagatagtaagacaccttattgccagcaaaaaat	600				
Db	541	ATTATTAAAGAGGGGCACATTGGGGCTTAGATAGTAAGACCTCTTATGCCAGCAAAAAAT	600				
Qy	601	agccaatttttaacgtctatcgcagcggctggtgagggcgggtggttttgaagcacttgct	660				
Db	601	AGCCATTTTTTACAGTCTATCGACGGCTGGTGAGCGGGTGGTTTTGAAGCACTTTGTT	660				
Qy	661	attgcaactcaccgacacggtaaagagagaccaaaattcattccaggcgaataaaattaaaa	720				
Db	661	ATTGCAACTACCAGACGGTAAAGAGACCANAATTCATTCCGAGGCAANATAAATTAAAA	720				
Qy	721	cataatattcggcgtataaccggctttgaaaaatcgctacgacttttaccocaaattccgcac	780				
Db	721	CATAATATTCGGCGTATAACCGGCTTCGAAAATCGCTACGACTTTACCCAAAATCCGCAC	780				
Qy	781	agaatgctcctcggaggaatctccttttaattgtggaagatacttgcccaacattagattgt	840				
Db	781	AGAATGCTCCTGGAGGATCTCCTTTAATTGTGGAAGATACTTGCCCAACATTAGATTGT	840				
Qy	841	actcctcgtgcaagggttaagtgaaccgcgataaatttcccagtgagaaacatttccggaa	900				
Db	841	ACTCCTCGTGAAGGGTTAAGTTGAACCGCGATAAATTTCCAGTGAGAACATTTCGGGAA	900				
Qy	901	tatacgcctgaagagcgcaaacagcttgagcagatttccttattcgcactgagcagctctca	960				
Db	901	TATACGCTGAAGACGCAACAGCTTGAGCAGATTCTCTTATCGCACATGAGCAGCTCTCA	960				
Qy	961	gcccaagaatatataccggtaaaagatcgcatgtgccacaaaccccttttagatacaaaagtaat	1020				
Db	961	GCCCAAGAATAATACCGGTAAGATCGCATGTGCACCAAAACCCCTTTAGATTACAAGAGTAAT	1020				
Qy	1021	tctgtttttatgaagtttggctatcatcacttcaactcgtctcattcttcttggcgcaattctta	1080				

Db	1021	TCGTGTTTTATGAAGTTTGGCTATCACCTTCACATCGTCTCATATATCTTGGCGCAATCTTA	1080				
Qy	1081	gaagatacaaaaaacacgctacgatatccgtgatgcgaacgccagcttactatacaaaa	1140				
Db	1081	GAAGATACAAAAACACGCTACGATATCCGTGATATGCAAAACGCCAGCTTACTATACAAA	1140				
Qy	1141	gaagataaacttatcaacttaggaactatggttatgaaggggataatattttatagatggc	1200				
Db	1141	GACGATATTAACTTATCACCTTAGGAATATGTTTATGAAGGGGATAATATTTTAGATGGC	1200				
Qy	1201	ttagtgtccaagccaagatcccttattgggttgcgctatagccaatgtgaagttttttgat	1260				
Db	1201	TTAGTGTTCAGCCCAAGATCCCTTATGGGTGGCTATAGCCATGTGAAGTTTTTTGAT	1260				
Qy	1261	gaacgtcaccacaacacgtctgtttaggtattcacctataataataaacccagagaataatcgc	1320				
Db	1261	GAACGTCAACCAAAACGTCGTTTAGGATTACCTTATAATATAAAACCAAGATAAATCGC	1320				
Qy	1321	tggttgatagcattaaactcagtcggaataaaacaaagatatattgaactatatagcggcta	1380				
Db	1321	TGGTTGATAGCATTAACCTCAGTCGGATAAACAAAGATATTGAACATATATAGCCGGCTA	1380				
Qy	1381	catcgcttgctttagcagattatccctgtgtagataaaaaattgcccgagacttttgat	1440				
Db	1381	CATCGCTTGCACTTAGGATTATCCTGTGATATAAATAATTCGCCGCCGACATTTGGAT	1440				
Qy	1441	aaactcttgggtctatgtatcgaactgagcgttaataattaccaagaaaaagcactcgtcatt	1500				
Db	1441	AAATCTTGGTCTATGTATCGAACTGAGCGTAATAATTAACCAAGAAAAGCATCGTGTCAAT	1500				
Qy	1501	catttagaatttgataaagcgttaaatgctggtcgaagcgtattttaaccacaaacccacaaa	1560				
Db	1501	CATTTAGAATTGATAAAGCGCTAAATGCTGGTCAAGCGGTATTTAACCAAAACCCACAAA	1560				
Qy	1561	ctgaatttagggttgggtcttggatcgatttaattcgtcttaggtatcattggggatagact	1620				
Db	1561	CTGAATTTAGGGTTGGGCTTTGATCGATTAAATTCGCTTATGGATCATGGGGATAGACT	1620				
Qy	1621	gcccaatataccaaaggcgggttataccagctaccgcggttagagggcgttttagataatcca	1680				
Db	1621	GCCCAATATACCAAGGCGGTTATACCAAGCTACCGCGGTAGAGGGCGGTTTAGATAATCCA	1680				
Qy	1681	tattttatcgcgcgcgataccagcagtatgaaacggtatctcttgttaataatacacgc	1740				
Db	1681	TATATTTATCGCCGCGATCCACGAGTATTGAAACGGTATCTTTGTGTAATAATACACGC	1740				
Qy	1741	ggcgacatctttaaaactgtgaaccgcgtaaaaattaaaggcgatagccattttgttagcttc	1800				
Db	1741	GGCGACATCTTAAACTGTGAACCGCGTAAATTAAGGCGGATAGCCATTTTGTAGCTTC	1800				
Qy	1801	cgcgacttagtgaacgaggtatggtgattgggattaggggtgcgttttgcatacaaat	1860				
Db	1801	CGCGATCTAGTGATAAGCGAGTATGGATTGGGATTAGGGGTGCGGTTTGTGATCAACAT	1860				
Qy	1861	cgatttaaatcgtgatcgtggacacttagccgaactatcgaaattggtcttgggaat	1920				
Db	1861	CGATTAAATCTGATGATCCGTGGACACTTAGCCGAACCTTATCGAAATTTGCTTTGGAAT	1920				
Qy	1921	gggtgggattacgcttaaaaccaacagagtttgcgtcttcttcttcgcatttcaaacggt	1980				
Db	1921	GGTGGGATTACGCTTAAACCAACAGAGTTTGTATCGCTTCTTATCGCATTTCAACCGGT	1980				
Qy	1981	tttagagtcgctgcatctctatgaactttatggtaaacgtgatcatatatgggcttaagat	2040				
Db	1981	TTTAGAGTCCTGCATTTCTATGAACCTTATGTTAAACGTGATCATATTGGGCTTAAAGAT	2040				
Qy	2041	aacgaatatgtgcaacgcgcgcaacgtagccacagtttagagccagaaaaatcgactaat	2100				
Db	2041	AACGAATATGTGCAACGCGCAACAGTAGCCAGTACCGAGCCAGAAAAATCGACTAAT	2100				
Qy	2101	catgagattggagttagcttttaaagggtcaatttgggttaccttgcgttgagctatttcggt	2160				

```
Db 2101 CATGAGATTGGATTAGCTTTAAAGGTCAAATTGGTTACCTTGA-----ATTTCGGT 2152
Qy 2161 aataactataaaaatatgattcgacagcatgtaaaagaataatacaaaaatcacactgt 2220
Pt 2161 aataactataaaaatatgattcgacagcatgtaaaagaataatacaaaaatcacactgt 2220
Pt 2161 aataactataaaaatatgattcgacagcatgtaaaagaataatacaaaaatcacactgt 2220
Ps 2153 AATAACTATAAAAATATGATTGCGACAGCATGTAAAGCAATAATACAAAATCACACTGT 2212
Cc 2221 ttctataactaccataaattatcaagatgtagcactaaacgggataaaattttagtcgctaaa 2280
Cc 2221 ttctataactaccataaattatcaagatgtagcactaaacgggataaaattttagtcgctaaa 2280
Cc 2213 TTCTATAACTACCAATTAATTAAGATAGACACTAAACGGGATAAAATTTAGTCGCTAAA 2272
Cc 2281 ttgacttacacgggtattttatctatgctccagatggtttttatttaccatggttcatt 2340
Cc 2281 ttgacttacacgggtattttatctatgctccagatggtttttatttaccatggttcatt 2340
Cc 2273 TTTGACTTACACGGTATTTATCTATGCTGCAGATGTTTTTATTCATCATGTTGCTTAT 2332
Cc 2273 TTTGACTTACACGGTATTTATCTATGCTGCAGATGTTTTTATTCATCATGTTGCTTAT 2332
Cc 2341 aacggtgtaaagtaaaagagcggaaactaacccgactcaagactcgatagcgttaaacgat 2400
Cc 2341 aacggtgtaaagtaaaagagcggaaactaacccgactcaagactcgatagcgttaaacgat 2400
Cc 2333 AACCGGTAAAGTAAAGAGCGGAAACTAACCGACTCAAGACTCGATAGGTTAAACGAT 2392
Cc 2401 cctattctagatcgattcagccagcagcgtatgtgcttgattggtggtacgatacccca 2460
Cc 2401 cctattctagatcgattcagccagcagcgtatgtgcttgattggtggtacgatacccca 2460
Cc 2393 CCTATTCTAGATGCGATTACGCCAGCAGCGCTATGCTTGGATTGCGCTACGATCACCCA 2452
Cc 2461 gaagaaaaatgggaattggcattactaccacattctctaaagcgaacgacgagtag 2520
Cc 2461 gaagaaaaatgggaattggcattactaccacattctctaaagcgaacgacgagtag 2520
Cc 2453 GAAGAAAAATGGGAATTTGGCATTACTACCACCTATTCTAAAGCCAAACGCCGATGAG 2512
Cc 2521 gtggcaggcacacgctcatcagcgatcacatcgctgtgattagtggtggcaaacctgaccggt 2580
Cc 2521 gtggcaggcacacgctcatcagcgatcacatcgctgtgattagtggtggcaaacctgaccggt 2580
Cc 2513 GTGGCAGGCACACGTCATCAGCGGATACATCGCGTGTGATTAGTGGCAAACTGACCGGTT 2572
Cc 2581 cttggtcacccatgatattaccggtttacatcaattataaaactacaccccttaacgtggag 2640
Cc 2581 cttggtcacccatgatattaccggtttacatcaattataaaactacaccccttaacgtggag 2640
Cc 2573 CTTGGTACACCATGATATTACCGGTTACATCAATTAATAAACTACACCTTACGTTGAG 2632
Cc 2641 gaattataatgtaactaatgtaaaattcccaatttgggaatcagtcgccaatcccggtg 2700
Cc 2641 gaattataatgtaactaatgtaaaattcccaatttgggaatcagtcgccaatcccggtg 2700
Cc 2633 GAATTTAATATGTGACTAATCGTAAATATCCACTTGGGAATCAGTGGCCCAATCCGGTG 2692
Cc 2701 tgaatgcagtaaaacaaagacggggtagcaattacactcgatttggcgctccggggagaa 2760
Cc 2701 tgaatgcagtaaaacaaagacggggtagcaattacactcgatttggcgctccggggagaa 2760
Cc 2693 TGAATGCAGTAAACCAAGACCGGGGTAGCAATTAACATCTCGATTGCGCTCCGGGGAGAA 2752
Cc 2761 atttcagtttagcatttgaaatgaagtttttag 2792
Cc 2761 atttcagtttagcatttgaaatgaagtttttag 2792
Cc 2753 ATTTTCAGTTTAGCATTTGAAATGAAGTTTAG 2784
Cc 2753 ATTTTCAGTTTAGCATTTGAAATGAAGTTTAG 2784

RESULT 2
T67241
ID T67241 standard; DNA; 3023 BP.
AC T67241;
DE 08-SEP-1997 (first entry)
DT Pasteurella haemolytica transferrin binding protein tbpa gene.
KW Transferrin binding protein; tbpa; receptor; vaccine;
KW Pasteurellosis; ss.
OS Pasteurella haemolytica serotype A1 strain H196.
FH Key
FT cds
FT 13
FT /*tag= a
FT /codon_start= 13
FT misc_difference 1115
FT /*tag= b
FT /note= "base 1115 is given as s in the
FT specification"
PN W09720934-A1.
PD 12-JUN-1997.
PF 29-NOV-1996; CA0791.
PR 01-DEC-1995; US-008569.
PR 01-DEC-1995; CA-164274.
PA (LORY/) LO B Y C.
PA (POT/) POTTER A A.
PA (SCH/) SCHRYVERS A B.
PI LO RYC, Potter AA, Schryvers AB;
```

```
DR WPI; 97-319780/29.
PT Pasteurella haemolytica transferrin binding proteins - used in
PT vaccines for prophylaxis and treatment of infection caused by
PT Pasteurella spp.
Pt Pasteurella spp.
Ps Example 1; Fig 3; 130pp; English.
Cc The tbpB (T67240) and tbpA (T67241) genes of Pasteurella
Cc haemolytica serotype A1 strain H196 were sequenced from recombinant
Cc clones obtd. from a P. haemolytica A1 gene library using a probe
Cc produced by PCR amplification (see also T67237-39). The genes are
Cc organised in an operon arrangement of tbpB-tbpa, the tbpB gene being
Cc preceded by a putative promoter and regulatory sequences. The two
Cc genes appear to be coordinately transcribed. Full-length coding
Cc sequences (T67235 and T67236) have also been obtd. from amino acid
Cc sequences for TbpA (W18061) and TbpB (W18062) have been deduced.
Cc Sequence 3023 BP; 890 A; 616 C; 667 G; 849 T;
SQ

Query Match 48.0%; Score 1339.8; DB 1; Length 3023;
Best Local Similarity 91.0%; Pred. No. 0;
Matches 1610; Conservative 0; Mismatches 97; Indels 62; Gaps 15;

Qy 1 atgataatgaataatcatcatttcgctatttccacttgccttaacacagtgatttgcgt 60
Db 13 ATGATAATGAATAATCATCATTTTCGCTATTTCACCTTGTGCTTAACAGTGTATTGCT 72
Qy 61 cttctcatcatcagctgctgcgactgaaataaaaaatacgaataacacgatcta 120
Db 73 CTTTCTCATTCATACGGTGTGCGACITGAATAAATAAATCGAAGAAATAACGATCTA 132
Qy 121 gctgttctggtatgaatttattgacagagagccattatgctcacgaacgctcaaaacgaa 180
Db 133 GCTGTTCTGGATGAAGTTATTGTGACAGAGAGGCCATTATGCTCACGAACGTCANACGAA 192
Qy 181 gtaactggtctggggaaagtagtgaaaaattatcacgaaatagtgaaaaatcacaattctt 240
Db 193 GTAACCTGGCTTGGGGAAGTAGTGAAAAATTTATCACGAAATAGTAAAAATCAAAATCTT 252
Qy 241 ggtattcgtgatttaacctcgtatgcctgacctgatttcggtggtgggaaagtcgaggt 300
Db 253 GGTATTCTGATTAATTAACCTGCTATGACCTCGTATTTCGGTGTGGTGAACAGGTGCGGGT 312
Qy 301 gcaagtgtggtcgtatgccattcggaggtgtagataaaaacccgtgctcagcttacttctgtat 360
Db 313 GCAAGTAGTGGCTATGCCATTCCAGAGGTGTAGATAAAAACCCGTGTCAGCTTACTTGTGAT 372
Qy 361 gggctacacaagcgcacagttatcatcagctaggttcagatgctaatggtggtgcaatt 420
Db 373 GGGCTACACACGCGCACAGTTATCATACGCT-GGTTTCAGATGCTAATGGTGTGCAATT 431
Qy 421 aatgagattgattgaaacacattcgttcaattgagtgaaagcaaggaagtcctgcg 480
Db 432 AATGAGATTGAGTAAGAAACATTCGTTCAATTAGTAAAGCAAGGAGCAAGTCTCGCG 491
Qy 481 gaattggtcctggtgcgcatggtggtgctattggttttcgtactaaagatgcgcagat 540
Db 492 GAATATGCTCTGGTGCATGTTGGTGTGCTATTGTTTTCGTACTAAGATCGCCAGGAT 551
Qy 541 attattaagaggggcagcattgggggttagtagtagtaagacctcttatgcagcaaaaat 600
Db 552 ATTATTAAAGAGGGGCAGCATTTGGGGCTTACATAGTAAGACCTCTTATGCCAGCAAAAAT 611
Qy 601 agccatttttaccagtctatcgacgctggtggtgagcggtgttttgaagcacttctt 660
Db 612 AGCCA-TTTTTTACAGTCTATCGCAGCGGTGGTTCAGCGGGTGGTTTTTCAAGCACTTGT 670
Qy 661 attgcaactcacgcacacggtgaaagagaccaaataattcctccagggcaataataataaa 720
Db 671 ATTGCAACTCACGACACGGTAAGAGACCAAAATTCATTCCGAGGCAAAATCAATT---A 727
Qy 721 cataatattcggogtataaccggcttggaaatcgctacgactttaccacaaatccgcac 780
Db 728 CATATTATTGGCGGTATACCGGGCTTTCAAAATCGCTACGACTTTAGCC-AATTCCGCAC 786
```

QY 781 agaagctcctgagggatctcctttaaattgtggaagatacttgcccaacattagattgt 840
 DB 787 AGAATGCTCTCTGGAGGATCTTTTATTTGTGGAAGATACTTGCCTCAACATTAGATTGT 846
 QY 841 actcctgctgaaggggttaagttgaacccgcgataatttccacagtgagaacatttccggaa 900
 DB 847 ACTCCTCGTCAAGGGTTAAGTTGAA-CGCGATAAATTTCCAGTGCAGAACATTTCCGAAT 905
 QY 901 tatacgctgaagagcgcaaacagcttgagcagattccttatcgcactgagagctctca 960
 DB 906 ATACGCTGGAAGAGGCGCAACAGCTTG---AGATTCTTTATCGCACTGCTCAGC- 961
 QY 961 gcccaagaataacacggtgaagacgcgattgacccaaacaccttagattacaagagtaat 1020
 DB 962 -CCAAAGAATAAACCGGTAAAGATCGGATTTGCACCAACCCCTTTAGATTACAAGAGTAAT 1020
 QY 1021 tctgtttttatgaagtttggtctatcaacttcaactcgctcctcatttcttcttggcgca-atctt 1079
 DB 1021 TCTGTTTTTATGAAGTTTGGCTATCACTTTTACCTCGTCTCATTATCTTGGCGCATCTCAC 1080
 QY 1080 agaagatacaaaaaacacgcgtacgatatcctgtgatagcacaacccagcttactatacaaa 1139
 DB 1081 AAGATGATACAAAACACGACGATATCCGTCATSTGCAAAACGCCAGCTTACTATACAAA 1140
 QY 1140 agacgataattcaactatcacttaggaactatgtttatgaaggggataataatttttagatg 1199
 DB 1141 AGACGATATTATCACTTTGGAACTATGTTTATCAAGGGGATATTA-TTTAGATGG 1199
 QY 1200 cttagtttcaagccaagggatcccttattggttgcgctatagccatggaagtttttga 1259
 DB 1200 CTTAGTGTCAAGCCAGGATCCCTTATGGGTTGCCGATATGCCATG- 1246
 QY 1260 tgaacgtacacacaaacgctgttaggattacactataataataaacagagaataatcg 1319
 DB 1247 TGAACGTCACCAACAGCTGCTTTAGGATTCACCTATAAATAATAAACAGAGAATAATCG 1306
 QY 1320 ctggttgatagcattaaactcag----- 1349
 DB 1307 CTGTTGGATAGCATTAACTCGTCGGTAGCTGCTTTCGCTCTGCTCTGAG 1366
 QY 1350 taacaagataattgaactatataagccggctacatcgcttgcattgttagcattatcctgt 1409
 DB 1367 TAACAAGATATTGAACATATATACCGGCTACATCGCTTGCATTGTAGCGATTATCCTGT 1426
 QY 1410 ggtagataaaaaattgcgcgcgcgactttggataaaatcttggctatgtatgcgaactgagcg 1469
 DB 1427 GGTAGATAAAAATTTCGCGCCGACCTTTGGATAAATC-TGCTTATGTATCGAACTGAGCG 1485
 QY 1470 taataattaccaagaagaagcatcgttcattcatttagaatttgataaagcgctaaatgc 1529
 DB 1486 TAATAATTACCAAGAAAGCATCGTGTCATTTTCAATTTTGAATTTGATAAGCGCTTAAATGC 1545
 QY 1530 tggtaagggcgtatttaaccacaccccaactgaatttagggttgggctttgctgatt 1589
 DB 1546 TGGTCAGGCGTATTTAGCAAAACCCACAACTGAATTTAGGCTTTGGGCTTTGAATCGAT 1605
 QY 1590 taattcgttatgatacatggttgatagctgacgccaatataccaaagg-oggtttatacca 1648
 DB 1606 TAAT--CGCTTATGATCATGGGGATATGACTTGCCCAATATACCAAAAGCCGGTTATACCA 1663
 QY 1649 gctaccggttagggcggttttagataatccatattattatccgcgcatccacgagta 1708
 DB 1664 GCTACCGCGG-AGAGGGGCTTTAGATAATCATATATTTATCGCGCGGATCCACGCACTA 1722
 QY 1709 ttgaacgggtatcttctgtgtaataaca 1737
 DB 1723 TTGAACGGTATCTTGTGTATAATACA 1751

RESULT 3
 ID T38071
 AC T38071; standard; DNA; 2826 BP.

DT 19-DEC-1996 (first entry)
 DE Transferrin binding protein 1 gene.
 KW Transferrin binding protein 1; Tbp1; vaccine; antibody; diagnosis;
 OS swine pleuropneumonia; ds.
 FH Actinobacillus pleuropneumoniae strain 1371 serotype 1.
 Key Location/Qualifiers
 rbs 1..4
 FT /*tag= a
 FT cds 11..2806
 FT signal_peptide 11..76
 FT /*tag= c
 FT mat_peptide 77..2803
 FT /*tag= d
 EP-733708-A2.
 PD 25-SEP-1996.
 PF 21-MAR-1996; 870033.
 PR 24-MAR-1995; ES-000592.
 PA (HIPR-) LAB HIPRA SA.
 PI Daban M, Espuna E, Medrano A, Querol E;
 DR WPI; 96-427056/43.
 DR P-PSDB; W04867.
 PT Actinobacillus pleuropneumoniae transferrin binding protein 1 - for
 production of antibodies useful diagnostically and in universal
 PT vaccine against porcine pleuropneumonia
 PS Claim 3; Page 10-17; 22pp; English.
 CC Recombinant plasmid pMAB:tr1 (CECT 4548) includes a sequence
 CC (T38071) coding for the transferrin binding protein 1 (Tbp1)
 CC of Actinobacillus pleuropneumoniae (App) strain 1370, derived
 CC from strain Hpn-1 (ATCC 27088) of serotype 1. It was obtd. by
 CC cloning a DNA fragment isolated from App genomic DNA into
 CC vector pUC119. The DNA fragment was identified using a probe
 CC based on a partial Tbp1 sequence identified using a Tbp2 gene
 CC fragment. The Tbp1 gene can be used to produce recombinant
 CC Tbp1 in transformed host cells, and can also be used to
 CC prepare diagnostic reagents. Tbp1 is useful for prodn. of
 CC antibodies and vaccines against porcine pleuropneumonia.
 CC Sequence 2826 BP; 925 A; 518 C; 596 G; 787 T;

Query Match 19.0%; Score 529.4; DB 1; Length 2826;
 Best Local Similarity 53.4%; Pred. No. 4.2e-142;
 Matches 1446; Conservative 0; Mismatches 1161; Indels 100; Gaps 12;
 QY 158 atgctcagaacgtcaaaaacgaagtaactggtctgggaaagtagtgaaaaattatcacg 217
 DB 126 AAGCACATATAAAAGAGACGAAAGTCACAGGCTTAGGAAAGTAGTTAAACACACAGATA 185
 QY 218 aatgagtaaaaaatcaaatcttgggtattcggtatttaactcgctatgacctggtattt 277
 DB 186 CTCTTAGTAGGACCAAGTGTTAGGAATACGAGATCTGACTCTGTTAGCACCCGGTATT 245
 QY 278 cgggtgggaaacaggtcgcggtgcaagtagtggtcattcgccattcgaggtgtagataaaa 337
 DB 246 CTGTCGTAGAACAGGAGAGAGGTGCGACTACAGCTACTCAATTCGGCGGTAGATCGTA 305
 QY 338 accggtcagcttacttgggttgatgggctacacaaagcgacagttatcatcacgtaggtt 397
 DB 306 ATCGTGTGGGCTTGGCATTAGACGGTTTGGCACAGATTCAATCTATGTAAAGCCAATATT 365
 QY 398 cagatgctaattggtgcaatgaagattgagattgaaacattcggttcaattgagt 457
 DB 366 CAGCTTCTCAAGCGGTGCCATTAAAGAAATAGAAATACGAAATCTCGGTTGATCCAAA 425
 QY 458 taagcaaaagagcaagttctgcggaatatggtcctggtgcgcatggtggtgctattggtt 517
 DB 426 TTAGTAAAGGGCTAGTTCTCTGAGTTTGGTAGTGCCTCACTAGGCGGTTCCGGTCAAT 485
 QY 518 ttcgtaataaagatgcgacaggtattattaaagagggcgacgattggggttagatagta 577
 DB 486 TCCGTACCAAGAGGTAAAGCGACATTTAAGCCAGGCAATCTTTGGGGATATTAGATACCA 545
 QY 578 agacctcttatgcagcaaaaaatagccattttttacagttctatcgacgcggtggtgagg 637


```
Db 3386 ATATTACAAAATAAGAACAAAGCGGCATCATTTGACAAAGCAGTGTAAAGTCTTAATC 3445
Qy 1352 aacaagatatgaactatatagccggctacatcgcttgcattgtagcattatctctgtg 1411
Db 3446 AACAAAACATCATCTTGACAGATTATATGCAACATACGCAATGCGAGTCTTTATCCTAATC 3505
Qy 1412 tagataaaattgcgcgcgactttggataaattgttctatgtatgcgaactagcgta 1471
Db 3506 CAAGTAAGAAATGGCGCCCAACAGCTGATAAACCTTATTCATATCATCATCTTGATAGAA 3565
Qy 1472 ataattaccaagaagaagcatgctgctcattcttagaatttgataaagcgcgtaaatgctg 1531
Db 3566 ATGTTTATAAGAAAACATATATGTTGCAATTCGAATTTAGAGAAAATAATTCACAAA 3625
Qy 1532 gtcaaggcgtatttaaccacacccacaaactgaatttagg-----ttgggctttg 1582
Db 3626 ATTGGCTTACTCATCAAAATGCTCTCAATCTTGTTTGTGATGACTTTACTTCAGCGCTTC 3685
Qy 1583 atcgatttaattcgttatgatcatggtgggatagactgcccataatacacaagcgcggtt 1642
Db 3686 AGCATAAAGATTATTTAACTCGACGTGTACCGTACGGCAAGAGTAGTTTCAGAGAAAG 3745
Qy 1643 ataccagctaccggttagaggggttttagataaaccatatttatccgcgcgcatccac 1702
Db 3746 CTATGAACAAAGAAAGATGGTTACAAAAAACACCTTACTTATACCCAAAACCAACAG 3805
Qy 1703 gcagttatgaacggtatctttgtgtaataacacg-----cgggacatctttaaact 1756
Db 3806 TAGGTTTGTAGTACAAGATCATTTGTGATTATTAAGGTAACTCTCTAAATFACAGAGACT 3865
Qy 1757 gtgaacgcgttaaaattaaagcgatagcgaattttgttagcttcgc---gatctagtga 1813
Db 3866 GTAAAGTGCGGTTAATTAAGGGAAAAATTAATTAATTTTCGACGACGCAATAATATGCGAT 3925
Qy 1814 taacgagatgtgtgattggataggggtgcgttttgatcaacatcgatttaaatctg 1873
Db 3926 TAGGAAATAGCTTGATTTAGTTTAGTTTACGTTATGCGTATGACGTATCTCGCACAAAAGCTA 3985
Qy 1874 atgacctgtgacacttagccgaacttatcgaaatttgcttggaaattgggtgggattacgc 1933
Db 3986 ATGAATCAACTATTAGTGTGGTAAATTTAAATAATTTCTCTTGGAAATAGTGGTATTGTCA 4045
Qy 1934 ttaaccacacagagttgtatcgctttcttctatcgattcattcaaacggtttttagagcgctg 1993
Db 4046 TAAACCAACAGGAATGGCTTGATCTTCTTATCGCTTTTCTACTGGATTAGAAATCCTA 4105
Qy 1994 cattctatgaacttttggttaaacgtgatcatattggcgttaagataaacgaatatgtgc 2053
Db 4106 GTTTTGTGAAATGTATGGTTGGCGGTATGGTGGCAATAATA----- 4147
Qy 2054 aacgcgcgaacgttagccaccagtttagagccagagaaaaatcgactaatcatgagattggag 2113
Db 4148 ---CGGAGGTTTATGTAGGTAATTTAAAGCTGAACATCTCGTAACCAAGAGTTTGGTC 4204
Qy 2114 ttagctttaaggtcaatttggttacotttgatgtgagctatttccgttaataactataaaa 2173
Db 4205 TCGCTCTAAAAGGGGATTTTGGTAATATTGAGATCAGTCATCTTTAGTAATGCTTATCGAA 4264
Qy 2174 atatgttgacagcatgtaaa---agaataatacaaaaaatcacactgtttctataact 2230
Db 4265 ATCTATCGCTTGTCTGAGAACTTAATAAATGGAATGGAAGCGCAATTTAGAT 4324
Qy 2231 accataattcaagatgtagcactaaacgggataaaatttagtcgtcaaaatttgacttac 2290
Db 4325 ATCATATGCACAAAATGCAAAATTTAGTTGGCGTAATAATAAATGCGCAATTTAGATTTA 4384
Qy 2291 acggtatttatctatcgtccagatgggtttttattcatcagttgcttataacogtgtaa 2350
Db 4385 ATGTTTATGGAACGATTTCCCTACGGTGGTATGCAACATTTGCTTATAACCGAGTAA 4444
Qy 2351 aagtataaagcgggaaactaacgactcaagactcgatagcgtataaacgacccatttag 2410
Db 4445 AAGTTAAGATC---AAAAAATCAATGCTGGTTGGCTCCGTAAAGCAGTTATTTATTG 4501
```

```
Qy 2411 atcgatttcagcagcagcgtctatgtcttgattcggctacgacccagcaaaaaa 2470
Db 4502 ATGCCATTAGCCCGAGCGGTTATATCATCTTGGTTAGGCTATGATCATCCAAGTAATACTT 4561
Qy 2471 ggggaattggcattactaccacacattctctaaagccaaaaacgcgcgatgaggtgcaggca 2530
Db 4562 GGGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGTAGGAA 4621
Qy 2531 caactcatcagcgatcacatgcg-----gttgatttaggtggcaaacgtgacggttctt 2583
Db 4622 AACGTGATTTGGGTAAACAATTCAGGGATGTAAATCAACAGAAAACCTTACTCGGCAT 4681
Qy 2584 ggtacacccatgatattaccgg---ttacatcaattataaaaaacacaccttactgtggag 2640
Db 4682 GGCATATCTTAGATGTATCGGGTTATTACATGGCGAATAAAATATTATGCTTCGATTAG 4741
Qy 2641 gaatttataatgtgactaatcgttaaatattccacttgggaatcagtcgcgaatcccggtg 2700
Db 4742 GGATATATAATTTATTCAACTATCGTATGTTACTTGGGAAGCGGTGCTGCAAAACAGCAC 4801
Qy 2701 tgaatgcagtaaaccaagacgcg---ggtagcaattacactcgatttggcgctccggga 2757
Db 4802 AAGTGGCGGTCAATCAACATCAAAATGTTGGTAGCTATATCTGCTACGACGATCAGGAC 4861
Qy 2758 gaaatttcagttgacgatttgaagaagttaa 2791
Db 4862 GAAACTATACCTTAACATTAGAAATGAATCTTA 4895

RESULT 5
V21431
ID V21431 standard; cDNA; 5099 BP.
AC V21431;
DT 20-JUL-1998 (first entry)
DE H. influenzae strain PAK 12085 transferrin receptor operon.
KW tbp1; tbp2; vaccine; H. influenzae; antibody; diagnosis;
KW passive immunisation; ds.
OS Haemophilus influenzae.
FH Key Location/Qualifiers
FT CDS 160..2124
FT /tag= a
FT /product= Tbp1
FT 2152..4893
FT /tag= b
FT /product= Tbp2
PN US5708149-A.
PD 13-JAN-1998.
PF 07-JUN-1995; 487890.
PR 08-NOV-1994; US-337483.
PR 08-NOV-1993; US-148968.
PR 29-DEC-1993; US-175116.
PR 07-JUN-1995; US-487890.
PA (CONN-) CONNAUGHT LAB LTD.
PI Chong P, Gray-Owen S, Harkness R, Klein M, Loosmore S,
PI Mordin A, Schryvers A, Yang Y;
DR WPI: 98-100410/09.
DR P-PSDB: W53048, W53049.
PT Purification of recombinant Haemophilus transferrin-binding protein
PT - by solubilising inclusion bodies separated from cell lysate
PS Example 4; Fig 6; 261pp; English.
CC The H. influenzae transferrin receptor operon contains two genes (tbp1
CC and tbp2) under the transcriptional regulation of one promoter. The
CC proteins encoded from these genes can be expressed in a recombinant
CC host. The proteins can be used in vaccines against H. influenzae
CC infections or to produce antibodies for use in diagnosis or passive
CC immunisation.
SQ Sequence 5099 BP; 1834 A; 849 C; 960 G; 1456 T;

Query Match 12.2%; Score 340.8; DB 1; Length 5099;
Best Local Similarity 50.4%; Pred. No. 8.3e-88;
Matches 1347; Conservative 0; Mismatches 1152; Indels 175; Gaps 14;
```

Qy	173	aaaacgaagtaaacctggcttggggaaagtagtgaaaaaattatcacgaaatagtaaaaaac	232
Db	2342	ATAATGAAGTAACCTGGACTGGCAAAATATCAAAACAGTAGTGAAAGTATCAGCCGAGAAC	2401
Qy	233	aaattcttggtattcgtgatttaactcgtcatgaacctgggtatttcogtgggtggaacaag	232
Db	2402	AAGTATTAAATATTCGTGATCTAACACGCTATGATCCAGGCGATTTCAGTTGTAGAACAG	2461
Qy	293	gtcgggtgcaagtagtggtctatgccattcggaggtgtagataaaaaaccgtgcagcttac	352
Db	2462	GCCGTGTCGAAGTCTGGATATCTATTTCGTGGTATGACACAGAAATAGAGTTCCTTTAT	2521
Qy	353	tgttgatgggtctacacaagcgacagtatacatcagctagggtc-----	398
Db	2522	TAGTAGATGGTTTACCTCAACCAATCTATGTAGTGCAAGGCCCTTTAGTTGCTCGTT	2581
Qy	399	---agatgctaatgggtgcgaattaaatgagatgagatgaaaaatcogttccaattg	454
Db	2582	CAGGATATCTCGCACTGGTGCNATTAATGAATTAATGAATATGAATAATGAAGCCGCTG	2641
Qy	455	agttaacgaaggaagcaagttctggcggaataatgaagaggggcagcattggggttagata	514
Db	2642	AAATAAGCAAGGGGGGAGTCTCTCGAGTAGTGTAAATGGAGCACTAGCTGTTCTGTAA	2701
Qy	515	gtttctactaaagatggcaggataattataaagaggggcagcattggggttagata	574
Db	2702	CATTTCAAAGCAATCAGCAGCGATATCTTAGAGGAGACAAATCATGCGGGAATTCAAA	2761
Qy	575	gtaagacctcttatgcagcaaaaaagcattttttcacgctcatcgcagcggctgggtg	634
Db	2762	CTAAAAATGCTTATTAACGCAAAAATAAGGCTTTACCCATCTTAGCTAGCTCGAA	2821
Qy	635	aggcgggtgtttgaagcacttggttatctgcaactcacccgacacaggttaaagagagcaaaa	694
Db	2822	AACAAGGGGATTTGACGGGGTCCGCATTTATCTACCAAGCAAAATCAATTAACCAACCAAG	2881
Qy	695	tctattccgaggaataataataaaacataatccgctggtataaacggctttgaaaaac	754
Db	2882	TCCATAAAGATGATTAAGGCGTACAAAGTTATCATCGATTATCGCC-----AAACC	2936
Qy	755	gctacgactttacccaaattccgcacagaaatgctcctcgaggagatcctctttaattggtg	814
Db	2937	AGAGGATCAATCTGCATCTTTGTGATGCAAGATGAGTGTCCAAAGCCAGATGATTATAA	2996
Qy	815	aagatacttcccaacattagattgactcctcctgcgaaggggttaagttgaaccgcgata	874
Db	2997	CAGTGTGTTTACCTTTC-----	3012
Qy	875	atttcccgatgagaacatttccggaataacgcctgaagcgcgaacagcttgagcaga	934
Db	3013	-----GCCAAAGGACCTCGATTT	3031
Qy	935	tctcttatcgactgagcagctctcagcccaagataatacgcggtaaagatcgcatgtcac	994
Db	3032	TATCTCCCAAGAGAAACCGTAAGCGTTTCAGATTATACGGGGCTAACCGGTATCAAAC	3091
Qy	995	caaacccttagattacaagagtaactctgtttttatgaagtttggtctatcaactcaact	1054
Db	3092	CTATCCAAATGAATATGAAGCCAGTCTTGTTTTTAAGAGAGGGGTATCATTTTTCTG	3151
Qy	1055	cgctcattatcttggcgcaactcttagaagatacaaaaacacgctacagatatccgtgata	1114
Db	3152	AACAACATTAATTTGGTGGTATTTTTGAATTCACACACAAAAATTTGATATCGCTGATA	3211
Qy	1115	tgaaacgccagcttactatacaaaagacgataattaacttataccttaggaactatgttt	1174
Db	3212	TGACATTTCCCGCTTATTTAAGATCAACAGAAAAACGGGATGATAGCAGTGGCTCTTTT	3271
Qy	1175	atgaaggggataatatttttagatggcttagtcttcaagccgaaggatcccttatgggttgc	1234
Db	3272	ATCCAAAGCAAGATTATGGTGCATCAACGATTATGGAGGTGCGC-----AGCGCTTA	3325

Qy	1235	gctatagccaatgtgaagttttttttagtagaagctcaccacaaacgctgttttaggattcaacct	1294
Db	3326	ACTATGCAAGTGGGCTTTATTTTCGATGAACACCATAGAAAAACACGCTGTAGGTATTGAAT	3385
Qy	1295	ataaataaaacacagagaataatcgtcgttggttagtagaattaaa---ctcagtgccgata	1351
Db	3386	ATATTTACGAAATATAGAACAAAGCGGCATCATTTGACAAAGCAGAGTGTTAAGTGTCTAATC	3445
Qy	1352	aaceagatatgaactatataagccggtcacatcgcttcgattgtagcattagcattatctctg	1411
Db	3446	AACAACAATCATACTTTCAGACGTTATATGCAACATACGCATTCGACGCTTTATCTCTAATC	3505
Qy	1412	tagataaaattgcgcgcgcgaactttggataaactttggtctctatgatacgaactgagcgta	1471
Db	3506	CAAGTAAGAATTGCGCCCAACACAGCTGATAAACCTTATTTCATCTATCTATCTTCATGAGAA	3565
Qy	1472	ataataccaagaaagacgtgctcattcatttagaatttgataaagcgcgtaaatgctg	1531
Db	3566	ATGTTTATNAAGAAACATAATATATGTTGCAATTGCAATTTTAGAGAAAAAATTCACACAA	3625
Qy	1532	gtcaaggcgtatttaaccaccccaaacactgaaattaggg-----tggggccttg	1582
Db	3626	ATTGGCTTACTCATCAAAATGCTTCATCTTCGTTTGTGATGACTTTACTTCACGCGTTC	3685
Qy	1583	atcgatttaattcgtcttagcatcggtggtatagactgcacataaccacaaaggcgtt	1642
Db	3686	AGCATAAAGATTATTTAACTCGACGTGTACCGCTACGGCAAGAGTATTTACAGAGAAAG	3745
Qy	1643	ataccagctaccggttagaggcgttagataatcaatatatttatccgcgcgatccac	1702
Db	3746	CTAATGAAACCAAGAAGAAATGTTTACAAAAACAACCTTACTTATACCCAAAAACCAACAG	3805
Qy	1703	gcagtatgaaacggtatcttctgtaataataacacg-----cgggcgacatcttaact	1756
Db	3806	TAGTTTTGTGTGTACAGAGTCATGTGTGATTTAAAGGTTAACTCTCTTAATTACAGAGACT	3865
Qy	1757	gtgaaccgcgtaaataaaggcgtagccatttggtagcttcgcg---gactagtga	1813
Db	3866	GTAAGTGGCGTTTAATTAAGSGMAAAATTAATTATTTTCGCAGCAGCAATAATATGCGAT	3925
Qy	1814	taecgagtagtggtgatttggttaggggtgctttgttgatacaacatcgatttaaatctg	1873
Db	3926	TAGGAAATACGTTGATTTAGGTTTAGGTATTCGGTATGACGTATCTCGCAACAAAGCTA	3985
Qy	1874	atgctccgtggcacacttagccaaactatcgaatgtgcttgaatgggtggattacgc	1933
Db	3986	ATGAATCAACTATTAGTGTGTGTAAATTTAAAAAATTTCTCTTGGAACTACGGTATTGTCA	4045
Qy	1934	ttaacacacagagttgttatcgcttctctatcgcatcttcaaacggtttttagagtgctg	1993
Db	4046	TAAACCAACGGAATGGCTTGATCTTCTTATCGCCTTCTACTTGGATTTAGAAATCCTA	4105
Qy	1994	cattctatgaactttatgttaaacgtgatacatattggttcctaaagataaogaatatgtgc	2053
Db	4106	GTTTGTGTGAATGTATGTTTGGCGGTATGTTGGCGGTATGTTGGCAATAATA-----	4147
Qy	2054	aacgcgcgaactagaccacagtagccacagtagccgttctatcgcatcttcaaacggtttttagagtgctg	2113
Db	4148	---CGAGAGTTTATGTAGGTAAATTTAAAGCCTTGAACATCTCGTAAACCAAGAGTTTGGTC	4204
Qy	2114	ttagctttaaggctcaatttggttaccttgatgtgagctatttcgcgtaataactataaaa	2173
Db	4205	TCGCTCTAAAAGGGGATTTTGGTAATATTGAGATCAGTCATTTTAGTAACTGCTTATCGAA	4264
Qy	2174	atagtattgcacagcatgtaaa---agaataatacaaaaatacnaactgtttctataact	2230
Db	4265	ATCTTATCGCCTTTGCTGAAGAACTTAATPAAAAATGGAACCTGGAAAGGCCAATATATGGAT	4324
Qy	2231	accataattcaagatgtagcactaaacgggataaaaatttagtcgctaaatttgaacttac	2290
Db	4325	ATCATATGCAAAATGCAAAATTAGTTGGCGTAAATATAACTGCGCAATTAGATTTTA	4384
Qy	2291	acggtattttatctatgctgccegatggttttttatctcatcagttgctttaaaccggttaa	2350

```

Db 4385 ATGGTTTATGGAACGATATTCCTACGGTGTGATGCAACATTTGCTTATAACCGAGTAA 4444
Qy 2351 aagtaaaagagcggaactcaacgactcaagactcgatagcgtaaacgactctattctag 2410
Db 4445 AAGTTAAAGATC--AAAATCATGCTGGTTGGCCCTCGGTAGCAGTATTTATTG 4501
Qy 2411 atgcgattcagccagcagctatgcttgattcggttagcgtacgacccagagaagaaaaa 2470
Db 4502 ATGCCATTACGCCAGCCGCTTATATCATTTAGGCTATGATCATCCAAAGTAATACTT 4561
Qy 2471 ggggaattggcattactaccacctattttaaagcaaaacgcgcgagtgagtgaggca 2530
Db 4562 GGGGAATTAATACATGTTACTCAATCAAAAGCAAAATCTCAAATGAATGCTAGGAA 4621
Qy 2531 cagctcatcagcgatacatcgc-----gttgattgagtgggcaaacactgaccggtctt 2583
Db 4622 AACGTGCATTGGGTACAAATTCAGGGGATGTAAATCAACAGAAATTTACTCGGGCAT 4681
Qy 2584 ggtacacccatgatattaccgg---ttacatcaattataaaactacaccttaacgtggag 2640
Db 4682 GGCATATCTTAGATGATATCGGTTATTACATGGCGAATAAAATATTATGCTCGATTAG 4741
Qy 2641 gaattataatgactaatcgtaaatattccacttgggaatcagtgcccaatcccggtg 2700
Db 4742 GGATATAATAATTATCAACTATCGCTATGTTACTTGGGAAGCGGTGCGTCAACACAGCAC 4801
Qy 2701 tgaatgcagtaaaacaaagaccgg---ggtagcaattacactcgatttggcgtccggga 2757
Db 4802 AAGGTGCGGTCAATCAACATCAAAATGTTGGTACTATACCTACGTCACGACATCAGGAC 4861
Qy 2758 gaaatttcagtttagcattggaatgaagttaa 2791
Db 4862 CAAACTATACCTTAACATTAGAAATGAATCTTA 4895

```

RESULT 6

```

ID Q94445 standard; DNA; 5099 BP.
AC Q94445;
DT 09-OCT-1996 (first entry)
DE Bacterial transferrin receptor operon (H. influenzae strain PAK 12085).
KW Tbp1; Tbp2; transferrin receptor operon; vaccine; antigen;
KW non-typable strain; Haemophilus influenzae; meningitis; ss.
OS Haemophilus influenzae strain PAK12085.
FH Location/Qualifiers
FT cds 160..2124
FT FT /*tag= a
FT FT /product= Tbp2
FT FT 2152..4896
FT FT /*tag= b
FT FT /product= Tbp1
FT FT 4891..4893
FT FT /*tag= c
FT FT /note= "no amino acid residue given in
FT FT corresponding protein"
PN W09513370-A1.
PD 18-MAY-1995.
PF 07-NOV-1994; CA0616.
PR 08-NOV-1993; US-148968.
PR 29-DEC-1993; US-173116.
PA (CONN-) CONNAUGHT LAB LTD.
PI Chong P, Gray-owen S, Harkness R, Klein M, Loosmore S;
PI Murdin A, Schryvers A, Yang Y;
DR WPI: 95-194089/25.
DR P-PSDB: R77889-90.
PT Nucleic acids encoding Haemophilus transferrin receptor - used to
PT develop prods for detection and in diagnosis, prevention and
PT treatment of Haemophilus infection.
PS Claim 12; Fig 6A-Q; 231pp; English.
CC The present DNA shows the transferrin receptor (Tfr) operon consisting of
CC two genes (Tbp1 and Tbp2) arranged in tandem and which are transcribed
CC from a single promoter, from the non-typable Haemophilus influenzae

```

```

CC strain PAK 12085. H. influenzae Tfr is iron- and/or haemin-regulated and
CC a putative fur-binding site has been identified upstream of tbp2.
CC Antibodies blocking this binding site may prevent bacterial growth.
CC Fragments of the Tfr (or its genes) are useful in vaccines to provide
CC protection against, e.g. bacterial meningitis. An advantage of using the
CC Tfr is that it shares homology with Tfr of other H. influenzae strains
CC including non-typable strains.
SQ Sequence 5099 BP; 1834 A; 854 C; 955 G; 1456 T;

```

```

Query Match 12.1%; Score 339.2; DB 1; Length 5099;
Best Local Similarity 50.3%; Pred. No. 2.4e-87;
Matches 1346; Conservative 0; Mismatches 1153; Indels 175; Gaps 14;

```

```

Qy 173 aaaaagaaagtaactgcttgggaaagtagtgaataattatcacgaatgagtaaaatc 232
Db 2342 ATAATGAAGTAACTGGCAATTTATCAAAACTAGTGAAGATATCAGCCGAGAAC 2401
Qy 233 aaattcttggtattcgtgatttaactcgtatgacctggtatttcggtggtgaaacaag 292
Db 2402 AAGTATTAAATATTCGTGATCTAACACGCTATGATCCAGGCATTTTCAGTTGAGACAA 2461
Qy 293 gtcggtgcaagtagtggctatccattccattgagtgtagataaaaaacggtgcagcttac 352
Db 2462 GCCGTGTGCAAGTCTGGATATTCTATCTGCTGATGGACAGAAATAGAGTTGCTTTAT 2521
Qy 353 ttgttgatggctaccacaagcgacagttatcaccgttagttc----- 398
Db 2522 TAGTAGATGGTTTACCTCAACGCAATCTTATGTAGTGAAGGCCCTTTAGTTGCTCGTT 2581
Qy 399 ---agatgctcaatggtggtgcaattaatgagtagtgatgaaacattcgttcaattg 454
Db 2582 CAGGATATTCTGGCAGCTGGTCAATTAATGAATTTGAATGAATGAATGAATGAATGAAT 2641
Qy 455 agttaagcaaaagagcaagttctcgtggaatgctcgtggtggtggtggtggtggtggtg 514
Db 2642 AAATAAGCAAGGGGGGAGTTTCTTCTGAGTATGTAATGAGCAGTACTAGTGGTCTGTAA 2701
Qy 515 gtttcgtactaaagatggtggtgcaattaatgagtagtgatgaaacattcgttcaattg 574
Db 2702 CATTTCAAGCAATATCAGCAGCCGATATCTTAGAAGGAGACAAATCATGCGGAATTCAAA 2761
Qy 575 gtaagacctcttgcagcaaaataatgagctatttttacagctctatcgagcggtggtg 634
Db 2762 CTAAAAATGCTTATTAAGCAAAATAAAGCCTTACCCTTCTTACCTGCTAGCTGGAA 2821
Qy 635 aggcgggtggttttgagcactgtttatgcaactcaccgacacggttaagagagaccacaa 694
Db 2822 AACAGCGGGATTTGACGGGTGCGCATTTATATCAACGAAATTCATTTGAACCCCAAG 2881
Qy 695 ttcatccggagcaataataataacataatttcggctgataacccggtttgaaaaatc 754
Db 2882 TCCATAAAGATGCTTAAACGGCTACAAAGTTATCATTCGATTAATCGCC-----AAACC 2936
Qy 755 gctacgactttaccacaaattccgcacagaatgctcctggaggatctccttttaattggtg 814
Db 2937 AGAGGATCAATCTGCATACATTGTGATGCAAGATGAGTGTCCAAAGCCAGATGATTATA 2996
Qy 815 aagataactgcccacacattagattgtactcctcgtgcaaggtttaagttgaacccgcata 874
Db 2997 CAGTTGTTTACCTTTC----- 3012
Qy 875 atttccagtgagaaacatttccgggaatatagcctgaagagcgcaaacagcttgagcaga 934
Db 3013 -----GCCAAACGACCTCGGATTT 3031
Qy 935 ttcttatcgactgagcagctctcagcccaagaataataccggttaagatcgcttgcac 994
Db 3032 TATCTCTCCCAAGAGAAACCGTAAAGCTTTTCAGATTTATACGGGGCTTAACGTTATCAAC 3091
Qy 995 caaaccttttagattacaagataattctgtttttatgaagtttggctatacactcaact 1054
Db 3092 CTAATCCAATGAATATGAAGCCAGTCTTGGTGTGTTTAAAGAGGAGGTATCATTTTTCTG 3151

```

```
Qy 1055 cgtctcattatcttggcgcgaatcttagaagatacaaaacacgcgtacgatccgtgata 1114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3152 AACACATATATATGGTGGTAATTTGNAATTCACACACAAATAATTTGATATCCGTGATA 3211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1115 tgcacacgcgcgtctactatacaaaagacgatattaacttatcacttaggaactatgttt 1174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3212 TGACATTTCCCGCTATATTTAAGATCAACAGAAACGGGATGATAGCAGTGGCTCTTTT 3271
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1175 atgaagggggataatttttagatggcttagtggttcaagcgaagatcccttatgggtgc 1234
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3272 ATCCAAAGCAAGAT-----TATGGTGCATATCAACGTATTTAGAGATGCCGAGCGCTTA 3325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1235 gctatagccatgtgaagttttttatgaacgtccaccacacgcgttttaggtatccact 1294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3326 ACTATGCAAGTGGCGTTATTTTCGATGAACACCATAGAAACACGCGTGTGATTGAAT 3395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1295 ataaataaaaccagagataaa----tcgctggttggtatagcattaaactcagtcgagata 1351
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3396 ATATTTACGAAATTAAGAACAAAGCGGCATCATTGACAAGCAGTGTTAAGTCTTAATC 3445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1352 aacaagatataaactatatacgccggtacatcgtctgcatgttagcgattatcctctg 1411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3446 AACAAAACATCATCTTGACAGTTATATGCAACATACGCATTGCAGTCTTTATCCTTAATC 3505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1412 tagataaaattgcgcgcgacttggataaactctggtctatctatcgaactgagcgta 1471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3506 CAAGTAAGAAATGGCGCCCAACACGATGATAAACCTTATTCATATCATCTCTGATAGAA 3565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1472 ataattcaaaagaaagcatcgtgtcattctattagaaatttgataaagcgcgtaaatctg 1531
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3566 ATGTTTATAAGAAACAAATATATGTGCAATTTGAATTTAGAGAAATAAATTCAACAAA 3625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1532 gtcaaggcgtatttaaccaaaacccacaaactgaatttaggg-----ttgggctttg 1582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3626 ATTGGCTTACTCATCAAAATGCTCTCAATCTTGTTTGTGACTTTACTTTCAGCGGCTC 3685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1583 atcgatttaactcgtctatgcatcgtgggatatagctcccaatataccaaaagcggtt 1642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3686 AGCATAAAGATTTATTAACTCGAGTGTTTACCGCTACGGCAAGAGTAGTTTCAGAGAAAG 3745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1643 ataccagctacccggtagaggcggttttagataatccatatattttatcgcgcgcatccac 1702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3746 CTAATGAACAAAGAAAGAAATGGTTACAAAACCAACCTTACTTTATACCAAAACCAACAG 3805
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1703 gcagtattgaacggtatctttgtgtaataacacg-----cggcgacatcttaact 1756
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3806 TAGGTTTGTAGTACAAAGATCATTTGTGATTATAAAGGTAACCTCTCTAAATTTACAGAGACT 3865
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1757 gtgaacgcgtataaaattaaagcgcgatagccattttgttagcttcgcg---gatctagtga 1813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3866 GTAAAGTGGGTTAATTAAAGGGAAATTAATTAATTTTCGGACGACGCAATTAATATGGCAT 3925
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1814 taacgagatgtgatttggttaggggtgcgtttttgatcaacatcgatttaaatctg 1873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3926 TAGGAAATACGTTGATTTAGGTTTAGTATTCGTTATGAGTATCTCGCACAAAGCTA 3985
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1874 atgatccgtggacacttagccgaacttatcgaaatttggcttggaaatgggtggaattacgc 1933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3986 ATGAATCAACTATTAGTGTTCGGTAAATTTAAAAATTTCTCTTGGAAATACTGGTATTGTCA 4045
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1934 ttaaaccaacagagtttgtatcgctttcttctcgcatttcaaacggttttagagtgctg 1993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4046 TAAACCAACCGAATGGCTTGATCTTCTTATCGCCTTTCTACTGGATTTAGAAATCCTA 4105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1994 cattctatgaactttatggttaaacgtagatcatatttgggctttaagataaacgaatatgtgc 2053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4106 GTTTTGTCTGAAATGTATGTTGGCGGTATGGTGGCAATAATA----- 4147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2054 aacgcgcgaacgtagccacacagtttagagcagcagaaaaaactaatcatcagatttgag 2113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4148 ---CGGAGGTTTATAGGTAATTTAAGCCTGAACATCTCGTAACCAAGAGATTGGTC 4204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Qy 2114 ttagctttaaaggtcaatttggttaccttgatgtgagctatttccgtaataactataaaa 2173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4205 TCGCTCTAAAAGGGGATTTTGGTAATATTGAGATCAGTCAATTTAGTAATGCTTATCGAA 4264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2174 atagtattgcagcatgtataa---agaataatacaaaaatcacactgtttcttaact 2230
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4265 ATCTTATCGCTTTTGTGGAAGAACTTAATAAAAAATGGAACCTGGAAGGCCAATATATGAT 4324
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2231 accataatatcaagatgtagcactaaaacgggataaaatttagtcgctaaattttgacttac 2290
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4325 ATCATATGCAAAAATGCAAAATTAGTTGCGCTAAATATAAATGCGCAATTTAGATTTTA 4384
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2291 acggtattttatcgtcgcagatggtttttattcatcagttgcttataaacggtgtaa 2350
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4385 ATGGTTTATGGAACGATTTCCTACGTTTGGTATGCAACATTTGCTTATAACCGAGTAA 4444
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2351 aagtaaaagcgcgaaactaaacgactcaagactcgtatagcgtataaacgactcctatttag 2410
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4445 AAGTTAAAGATC---AAAAATCAATGCTGTTTGGCTCCGTAAAGCAGTATTTATTATTG 4501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2411 atcgatttcagccagcagcgtctatgcttggattcggctacgatccccagaagaaaaat 2470
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4502 ATGCCATTTCAGCCAGCGCTTATATCATTTGTTTAGCTATGATCATCCCAAGTAATACTT 4561
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2471 gggaaattgcatctactaccacctattctaaagccaaaacgcgcgatgaggtggcagca 2530
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4562 GGGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATGTGTAGGAA 4621
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2531 caggtcatccgcgtacatcgc-----gttgatttagtggcgaactgaccggttctt 2583
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4622 AACGTGCTATTGGGTAAACAATTAAGGGATGTAAATCAACAAGAAATTTACTCGGSCAT 4681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2584 ggtacacccatgatattaccgg---ttacatcaattataaaaactacacattacgtggag 2640
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4682 GGCATATCTTAGATGTATCGGTTATTACATGCGCAATAAAAAATTTATGCTTCGATTAG 4741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2641 gaattataatgtgaactaatcgtataattccacttgggaatcagtcgcgaatcccggtg 2700
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4742 GGATATATAATTTATTCAACTATCGCTATGTTACTTGGGAAGCGGTGCGTCAACACGAC 4801
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2701 tgaatgcagtaaaccaagaccgg---ggtagcaattcacctcgatttggcgctcccgga 2757
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4802 AAGGTGGGGTCAATCAACATCAAAATGTTGTTAGTACTCGTACGACGATCAGGAC 4861
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2758 gaaatttcagtttagcatttgaagaagttaa 2791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4862 GAAACTATACCTTAACATTTAGAAATGAATTTCTA 4895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 7

T49504

ID T49504 standard; DNA; 5144 BP.

AC T49504;

DT 05-MAY-1998 (first entry)

DE Transferrin receptor gene sequence encoding Tbp1 and Tbp2.

KW Transferrin receptor; Haemophilus influenzae type b;

KW iron; human transferrin; iron source; antibody; bacterial growth;

OS Haemophilus influenzae.

FH key Location/Qualifiers

FT CDS 192..698

FT /*tag= a

FT /note= "encodes Tbp2 (W08966)"

FT 2135..4870

FT /*tag= b

FT /note= "encodes Tbp1 (W08965)"

FT W09640929-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; CA0399.

PR 17-MAY-1996; US-6495518.

PR 07-JUN-1995; US-483577.

PA (CONN-) CONNAUGHT LAB LTD.

PI Chong P, Gray-Owen S, Harkness RE, Klein MH, Loosmore SM,

Db 2805 AACAGGTGGATTGAAGGGGTCGCCATTTACACTCAACGAATTCGGAGGAACCCAG 2864
Qy 695 ttcatccgaggaataataataaataattcattcggtataacccgctttgaaatc 754
Db 2865 TCCATAAAGATGCATTAAAGCGGTACAAGTTATGAGCGATTCATCGCC----ACAACA 2920
Qy 755 gtaacgactttaccacaaatccgcacagaatgctcctcgtgagatcctcttaattgtgg 814
Db 2921 GATAAATCTTCAGGATACATTTGTGATACAAGGTGAGTG-----TCCAAATGGTG 2969
Qy 815 aaataacttgcacacattagattgtaactcctcgtgcaagggttaagttgaaccgcgata 874
Db 2970 ATGCAAGTGTGCAGC----- 2985
Qy 875 atttccagtgagacatttccggaataataacgcctgaagagcgcaaacagcgttgagcaga 934
Db 2986 -----CAACACCCTGCAAGT 3002
Qy 935 ttccctatgcactgagcagctctcagcccaagaataataccggtaaaagatcgcatgcaac 994
Db 3003 TATCCCCCAAGCGAAACCGTAAGCGTTTCAGATTATACGGGGGTAAACCGTATCAAAAC 3062
Qy 995 caaaccttttagattacaagagtaattcgtttttatgaagtttggtctatcaactcaact 1054
Db 3063 CTAATCCCAATGAATATGAAGCCAGCTCTGGTTTTTAAGAGAGGGGTATCATTTTTCTG 3122
Qy 1055 cgtctcattatcttgccgaactcttagaagatcacaaaaacagcgtcacatatccgtgata 1114
Db 3123 AACACACTATATGGTGGTATTTTGAATTCACACACAAAAATTTGATATCCGTGATA 3182
Qy 1115 tgcaacgcgcagcttactatacaaaagcagatattaacttatcacttaggaactatgttt 1174
Db 3183 TGACATTTCCCGCTTATTTAAGATCAACAGAAAAACGGGATGATAGAACTGSCCCTTTT 3242
Qy 1175 atgaaggggataaatttttagatgcttagtgttcaagcccaaggatcccttatgggtgc 1234
Db 3243 ATCCAAAGCAAGATTATGGTGTGATATACACGTATTGAGGATGGCCGA-----GGCGTTA 3296
Qy 1235 gctatagccatggaagttttttagtaacgtccacaaacgctgcttaggattcaact 1294
Db 3297 ACTATGCAAGTGGGTTTATTTGATGACACACCATAGAAAACAGCGGTGATGATTTGAAT 3356
Qy 1295 ataaataaaacagagaataatcgtggttgatagcattaaactc---agtcgggata 1351
Db 3357 ATATTTACGAAAAATAAGAAACAAAGCGGCATCATTTGACAAAACAGTGTAAAGTGCTAATC 3416
Qy 1352 aacaagatatgaactatatacggcgctacatcgtctgcatgttagcagattatcctgtgg 1411
Db 3417 AACAAAACATCATCTTGACATTTATGCGACATATGCGACATACGCATTTGCAGTCTTTATCCATAATC 3476
Qy 1412 tagataaaaaatgcccgcgacttgggataaactctgtctatgtatgaaactgagcgata 1471
Db 3477 CAAAGTAAGNATTCGCCCGCACACTTGATAAACCTTTATCATTATCGTCTGTATAGAA 3536
Qy 1472 ataatacaagaagaagcatcgttcattcatttagaatttgataaagcgtcaaatgctg 1531
Db 3537 ATGTTTATAAGAAAAACATAATATGTTGCAATTCGAATTTAGAGAAAAAATTTCAACAAA 3596
Qy 1532 gtaaggggtatttaacacaaaccccaaaactgaatttggg-----ttgggctttg 1582
Db 3597 ATTGCTTACTCATCAATTTGCTCAATCTCTGGTTTTGATGACNTTACTTCAGCGCTTC 3656
Qy 1583 atcgatttaactcgtttatggtatcgtgggataatagctgcccataatacacaagcggtt 1642
Db 3657 ACATAAAGATTTATTTACTCGACGTGTTACCGCTACGGCAAAATATTTATTCAGGGACAG 3716
Qy 1643 ataccagctaccggttagagggcgttttagataatccatatatttatcgccgcatccac 1702
Db 3717 TTGCTGGTAAACGAAGAAATGTTACGAAAAACACCTTACTTTACTCAAAACCAAAAG 3776
Qy 1703 gcagttgaaacggttatcttttgtgtaataataacacgggc-----gacatcttaaac 1756

Db 3777 TAGATTTGTAGGACAAGATCATTTCTAATATATAAAGTAGTCTCTTAATTACAGCGACT 3836
Qy 1757 gfgaacccgcgtaaataataaagcgatagccattttgttagcttccgc---gatctagta 1813
Db 3837 GTAAAGTCGGTTAATTAAGGGAAAAATTAATTTCCGACGACGCAATAATATATGGCAT 3896
Qy 1814 taagcgagtagtggtgatttgggatttaggggtgcttttgatcaacatcgatttaactcg 1873
Db 3897 TAGGAAAAATCATTTGATTAGTTTAGTATTTCGTATGACGTATCTCGTACAAAAGCTA 3956
Qy 1874 atgacctcgtgagacacttagcgaacattatgaataattggtccttgggaattggtggaatacgc 1933
Db 3957 ATGAATCAACTATTAGTCTTGTAAATTTAAAAATTTCTCTGGAAATCTGCTATTGTCA 4016
Qy 1934 ttaaaccaacagagttgtatcgtcttctatcgatttcaaacggtttttagagcgctg 1993
Db 4017 TAAAACCAACGGAATGGCTTGATCTTTCTTATCGCTTTTCTAGGATTTAGAAATCTCTA 4076
Qy 1984 cattctatgaactttatggttaacggtgatcatattggccttaaaagataaacgaatatgtgc 2053
Db 4077 GTTTTGTGTAATGATGTTGGGGTATGTTGGCAATAAATAGCGATGTTTATGTAGTA 4136
Qy 2054 aacgcgcgaacgtagccaccagtttagagcgagaaaaatcgaactaatcatgagattgag 2113
Db 4137 AA-----TTTAAGCCTGAAACATCTCGTAACCAAGAGTTTGGTC 4175
Qy 2114 tttagctttaaaggtcaatttggttacccttgatgagctatttccgttaactaataaaaa 2173
Db 4176 TCGCTCTAAAAGGGGATTTTGGTATATTGAGATCAGTCATTTTAGTAAATGCTTTATCGAA 4235
Qy 2174 atatga-----ttgcgacagcatgtaaaagaataatacaaaaaatcacactgtttctata 2227
Db 4236 ATCTATCGCCTTCTGCTGAAGACTTAGTAATAATGGAACACTACTGGAAGGCAATTAAG 4295
Qy 2228 actacataatattoagatgtagcaactaaacgggataaaatttagtgcataaattgact 2287
Db 4296 GATATCAATAATGCACAAAATAGTTGGCGTAATAATAACTCGCAATTAGATT 4355
Qy 2288 taccggtattttatctatgctgcagatggtttttatctatcagttgcttataaactg 2347
Db 4356 TTAATGGTTTATGGAACGTATTCCTACGGTGTGATGCAACATTTGCTTTATAACCGAG 4415
Qy 2348 taaagtataaagagcggaaactaacccgactcaagactcgatagcgtaaacgactcatttc 2407
Db 4416 TAAAGTTAAAGATC---AAAAATCAATGCTGGTTGCGCTCCGTAGACAGTTATTAT 4472
Qy 2408 tagatgcattcagccagcagctatgcttggtgattcgtgattcgttaoagatcacccagaagaa 2467
Db 4473 TTGATGCCATTACGCCACCGCTTATATCATTTGGTTTAGGCTATGATCATCAAGTAATA 4532
Qy 2468 aatggggaattggcattactaccacttatcctaaagcacaacccgcatgaggtggcag 2527
Db 4533 CTGGGGGAATTAATCAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTTGCTAG 4592
Qy 2528 gcacagctcatcaggtacatcgc-----gttgatttaggtggcaaacgactgacccgtt 2580
Db 4593 GACAACGTGCATTTGGTTAAACAATTCAGGAATGTAATAATCAACAGAAACTTACTCGGG 4652
Qy 2581 cttggtaacccatgatatattaccg-----ttacataaattataaaactacaccttactg 2637
Db 4653 CATGCAATCTTATAGATGATCGGTTTATACATGGCGAATAAAAAATATTATGCTTCGAT 4712
Qy 2638 gagaaattataagtgactaatcgttaaatccacttgggaatcaatgagcgcaactcg 2697
Db 4713 TAGGATATAATAATTATTCAACTATCGCTATGTTACTTTGGGAACGCGTGGCTCAACAG 4772
Qy 2698 gttgtaagtcagtaaaccaagacc---ggggtagcaaatcacactgatttggcgctcgg 2754
Db 4773 CACAAGTGGCGTCAATCAACATCAAAATGTTGGTAGCTATACTCGCTACGAGCATCAG 4832
Qy 2755 ggagaaatttcagtttagcatttgaatgaagtta 2791
Db 4833 GACGAACTATACCTTAACATTAGAAATGAAATCTTA 4869

RESULT 9
 T49502 T49502 standard; DNA; 5009 BP.
 AC T49502:
 DE 05-MAY-1998 (first entry)
 DE Transferrin receptor gene sequence encoding Tbp1 and Tbp2.
 KW Transferrin receptor; Haemophilus influenzae type b;
 KW iron; human transferrin; iron source; antibody; bacterial growth;
 KW vaccine; immunogenic truncated analogue; antigen; Tbp1; Tbp2; ss.
 OS Haemophilus influenzae.
 FH Key Location/Qualifiers
 FT CDS 121..2103
 FT /*tag= a
 FT /note= "encodes Tbp2 (W08962)"
 CDS 2117..4855
 FT /*tag= b
 FT /note= "encodes Tbp1 (W08961)"
 PN W09640929-A2.
 PD 19-DEC-1996.
 PF 07-JUN-1996; CA0399.
 PR 17-MAY-1996; US-649518.
 PR 07-JUN-1995; US-483577.
 PA (CONN-) CONNAUGHT LAB LTD.
 PI Chong P, Gray-Owen S, Harkness RE, Klein MH, Loosmore SM,
 DR WPI; 97-052329/05.
 DR P-PSDB; W08961-62.
 DR Tbp1 truncated transferrin receptor protein analogue, Tbp2 -
 PT used to induce protection against disease caused by transferrin
 PT producing pathogens, or as antigen to detect Haemophilus Tfr
 PT antibodies
 PS Claim 5; Fig 5A-Q; 228pp; English.
 CC The present sequence represents the transferrin receptor gene of
 CC Haemophilus influenzae type b, strain Minna. The bacterial transferrin
 CC receptor is composed of 2 chains, Tbp1 and Tbp2. H. influenzae is a
 CC non-encapsulated or non-typable bacterium responsible for a wide range
 CC of human diseases. Iron is an essential nutrient for the growth of these
 CC bacteria, and they can utilise human transferrin as a source of iron.
 CC Antibodies which block the access of the transferrin receptor to
 CC its iron source prevent bacterial growth. The transferrin receptor, or
 CC fragments, therefore, are good vaccine candidates. The full length Tbp2
 CC protein is produced in low amounts in Escherichia coli. However, the
 CC yield can be enhanced by truncation of the 3' end of the gene. An
 CC immunogenic composition comprising (or encoding) the immunogenic
 CC truncated analogue can be used to induce protection against a disease
 CC caused by a bacterial pathogen that produces the transferrin receptor.
 CC The immunogenic truncated analogue is also useful as an antigen in
 CC immunoassays for the detection of Haemophilus transferrin receptor
 CC antibodies, while the nucleic acid molecule can be used as a
 CC hybridisation probe for the detection of other transferrin receptor
 CC genes.
 SQ Sequence 5009 BP; 1809 A; 836 C; 934 G; 1430 T;

Query Match 11.9%; Score 331.2; DB 1; Length 5009;
 Best Local Similarity 50.5%; Pred. No. 4.7e-85;
 Matches 1354; Conservative 0; Mismatches 1133; Indels 193; Gaps 15;

Qy 173 aaacgaagtaactgcttgggaaagtagtgaaataattatcacgaaatgataaaatc 232
 Db 2307 ATAATGAAGTAAGTCTGCAAAATATCAAAACATAGTGAAGTATACGCGGAGAAC 2366
 Qy 233 aaattcttggtattgatttaactgctatgaccttggtatttcggtggtggaacaag 292
 Db 2367 AAGTATTAATATTCGTATCAACCGCTATGATCCAGGATTTTCAGTGTGAGAACAAG 2426
 Qy 293 gtcgcggtgcaagtgtggtgctatgccatcgaggtgtgatgataaaacccgtgtcagcttac 352
 Db 2427 CTCGCGGTGCAAGTCTCGATATTCATTCGTGGTATGCGACATAACGATTCGCTTTAT 2486
 Qy 353 ttgttgatgggtaccacgaagcgacagttatcatcacgctaggttc----- 398

Db 2487 TAGTAGATGGTTTACCTCAAAACGCAATCTTATGTAGTGCAAGCCCTTTAGTTCCTGCTT 2546
 Qy 399 ----agatgctaattggtgtaataatgagattgagattgataaaacattcgttcaattg 454
 Db 2547 CAGGATATTCGACACTGGTCAATTAATGAAATGAAATGAAATGAAAGCCGCTCG 2606
 Qy 455 agttaagcaagaggaagttctgctggaatatgctctggtgagcatggtggtgctattg 514
 Db 2607 AAATAAGCAAGGGGGGAGTCTTCTGATGTGTAATGAGGACACTAGCTGGTGTCTTAA 2666
 Qy 515 gttcttactaaaagatgctgcagagatattataaagggggcagcattgggggcttagata 574
 Db 2667 CATTTCAAAGCAATCAGACGCCGATATCTTAGNAGAGACAAATCATGCGGAATTCANA 2726
 Qy 575 gtaagacctcttatgccagcaaaaatagccattttttaccagtctatcgagcggtcgtg 634
 Db 2727 CTAAAAATGCTTATTCAAGCAAAAATAAGGCTTTTACCATTCTTTAGCTGTAGCAGAA 2786
 Qy 635 aggcgggtggttttgaaagcactgttattgcaactcaccgacacggttaagagagacaaaa 694
 Db 2787 AACAAAGGTGGATTGAAAGGGCTAGCCATTTACACTCAACGAAATTTCAATTGAAACCCAAG 2846
 Qy 695 ttcaatccgagggcaataataataaaacataatctcggcgataaacggcctttgaaatc 754
 Db 2847 TCCATAAAGATGCATTAAGGGCGTACAAAGTTATGATCATTAATCGCC-----ACAACA 2902
 Qy 755 gctacgactttaccocaaattccgcacagaatgctcctggaggatcctcctttaaattg 814
 Db 2903 GATAAACTTCAGGATACTTTGTGATACAAGGTGAGTG-----TCCAATAGGTG 2951
 Qy 815 aagatacttcccaacacattgagattgtactcctcgtgcaaggggttaagtgaacgcgcata 874
 Db 2952 ATGCAAGTGTGCAGC----- 2967
 Qy 875 attccccagtgagacaattcccggaatatacgcctgaagcgcgaacagcgttgagcaga 934
 Db 2968 -----CAAGCCACTCGGACTT 2984
 Qy 935 ttcttactgcactgagcagctctcagcccaagaataatcacgcgttaagatgcgacttcac 994
 Db 2985 TATCCACCCAAAGGAAACCGTAAGCGTTTCAGATTATACGGGGCTAACCGTATCAAAC 3044
 Qy 995 caaacctttagattacaagagtaattctgttttgaagtttggtctatcacctcaact 1054
 Db 3045 CTAATCCAATGAATATGAAGCCAGCTTGGTGTGTTTAAAGAGAGGGTATCATTTTCTG 3104
 Qy 1055 cgtctcattcttggcgcaatcttagaagatacaaaacacgcctacgataccgctgata 1114
 Db 3105 AACAAATATTATTTGGTGGTATTTTGAATTCACACACAAATAATTTGATATCGTGATA 3164
 Qy 1115 tgcaaacgcgcagcttactatatacaaaagacgatatataacttcatcacttaggaactatgtt 1174
 Db 3165 TGACATTTCCCGCTTATTTAAGCCCAACAGAAAGACGGGATGATAGTAGTCTCTTTT 3224
 Qy 1175 atgaaggggataaattatttagatggcttagtgttcaagccaaggaatcccttatgggtg 1234
 Db 3225 ATCCAATGCAAGATCATGGTGCATATCAACATATTGAGGATGGCA-----GAGCGGTTA 3278
 Qy 1235 gctatagccatgaaagtgtttttgatgaacgtcaccaacacgcgttttagattcacct 1294
 Db 3279 AATATGCAAGTGGCTTTATTTCCGATGNACACCATAGNAACACGCGTGTAGTATTGAAT 3338
 Qy 1295 ataaataaacaccagagaataatcgctggttggtatagcattaaactc---agtgcgggata 1351
 Db 3339 ATATTTACGAAATAAGAAACAAAGCGGCATCATTTACAAAAGCAGTGTAAAGTGTAAATC 3398
 Qy 1352 aacagatattgaactatagccgggtacatcgctgcattgagcattatcctgtg 1411
 Db 3399 AACAAAACATCATCTTGACAGTTATATGCGACATACGATTCGATTCCTTATCCTAAATC 3458
 Qy 1412 tagataaaattgcccgcgcacttgggataaattccttggtctatgtatcgaaactgagcga 1471

```
Db 3459 CAAGTAAGATTGCGCCCAACACATTGTGATAAACCTTATTATCATATATCGTTCTGTATAGAA 3518
Qy 1472 ataattaccaagaagcatcgtgcatcattagattgataaa----- 1518
Db 3519 ATGTTTATAAGAAAAACATAATATGTTGCAATGCAATTAGAGAAAAAAATTCACAAA 3578
Qy 1519 --gcgcataatgctggtcaagcgttatttaaccaaaccccaaacgaatgaggttgg 1576
Db 3579 ATTGGCTTACTCATCAAAATGCTTCAATCTTGTTGATGACATTTACTTCAGCGCTTC 3638
Qy 1577 gcttgatcgatttaattcgcgttatggtatcggtggtatgagtgcccaataaccagaag 1636
Db 3639 AGCAATAAGATTATTTAACTACGACGTGTTATGCGTACGCGCATAGTATTCACAGGAAC 3698
Qy 1637 ggggttataccagctaccgcggtagagggcgttttagataatccatattatcgccgcg 1696
Db 3699 CTGGTGAACCTGGTAAACCAAGAATGTTTGCATATCACACCTTACTTATACCCAAAC 3758
Qy 1697 atccacgcagttattgaacggttatcttctgtgtaataatacacgggc-----gacatct 1750
Db 3759 CAGAGCCATATTTTGCAGGACAAGATCATTTGTAATATATCAAGGTAGCTCCTCTAATTACA 3818
Qy 1751 taaactgtgaacgcggtataaataaagcgcgtagccatttggtagcttcgc---gac 1807
Db 3819 GAGACTGTAAAGTGGGTAAATTAAGGAAAAATTTATTATTCGCACGACGCAATAATA 3878
Qy 1808 tagtataagcagtagtattggtattggttaggggtgctgtttgatcaacatcgattta 1867
Db 3879 TGGCATTAGGGAATACGTTGATTTAGGTTTAGGTATTTCGGTATGACGTATCTCGTACAA 3938
Qy 1868 aatcgtatgctcgtggacactcgcgaacttatcgaaatttgcttctgggaatggtggga 1927
Db 3939 AAGCTAATGAATCAACTATTAGTGTGGTAAATTTAAAAAATTTCTCTTGGAACTACTGTA 3998
Qy 1928 ttacgcttaaacacagattgtatcgcttcttctatcgcatttcaacagggttttagag 1987
Db 3999 TTGTCATAAAACCAACGGAATGGCTTGATCTTTCTTATCGGCTTTTCTACTGGATTAGAA 4058
Qy 1988 tgcctgcatcttatgaactttatgtaaacgtagcatattgggtcttaagaatacgaat 2047
Db 4059 ATCTTAGTTTTCGAATATGATGTTGGCGGTATGGTGGCAAGNATCACGAGGTTTATG 4118
Qy 2048 atgtgcaacgcgcgcaacgtagccaccagtttagagccgagaaaaatcgactaatgaga 2107
Db 4119 TAGGTAAA-----TTTAAGCCCTGAACATCTCGTAACCAAGAGT 4157
Qy 2108 ttgagttagcttttaaaagtcattggtttacccttgatgtagctattcttcgtaataact 2167
Db 4158 TTGGTCTCGCTCTAAAGGGGATTTGGTAATATTAGATCAGTCATTTTAGTATGCTT 4217
Qy 2168 ataaaaatgatgtgcagacgatgtaa---agaataatacaaaaatcacactgtttct 2224
Db 4218 ATCGAAATCTTATCGCCCTTTCGTGAAGAACCTTAGTAAAAATGGAACCTGGAAGGGCAATT 4277
Qy 2225 ataactaccataataatccaagtagtagcactaaacgggataaaattagctcgtataattg 2284
Db 4278 ATGGATATCATAAATCACAAAATGCAAAATAGTTGGCGTAAATATATACTGCACAAATTAG 4337
Qy 2285 acttacaagggtatttatctatcgtgcagatggtttttattcatcagttgcttataacc 2344
Db 4338 ATTTTAAATGGTTTATGGAAACGATATTCCTACGGTGTGGTATGCAACATTTGCTTATAACC 4397
Qy 2345 gtgtaaaaagtaaaagcggaaactaacgcgactcaagcagctgatagcgaacgatccta 2404
Db 4398 AAGTAAAAAGTTAAAGATC---AAAAAATCAATGCTGGTTTAGCCTCCGTATAGCAGTTATT 4454
Qy 2405 ttctagatgcgattcagcgcagcacgctatgtgcttgggtattcgttcacgatcacccagaag 2464
Db 4455 TATTGTATGCCATTACGCCCGCGCTTATATCATTTGTTTGGTGTATGATCATCAAGTA 4514
Qy 2465 aaaaatggggaattggcattactaccacattcttaagagccaaaacccgcatgaggtgg 2524
Db 4515 ATACTTGGGGAATTAACAATGTTTACTCAATCAATCAAAAGCAAAATCTCAAAATGAATTGC 4574
```

```
Qy 2525 caggcacacgctcatcacggtatacatcgc-----gttgatttaggtggcaaacgacgcg 2577
Db 4575 TAGAAAAACGTGCATTAGGTAAACAATTCAAGGGGATGTAAAAATCAACAAGAAATTTACTC 4634
Qy 2578 gttcttggtacacccatgatattaccgg---ttacatcaattataaaaactacacacctac 2634
Db 4635 GGGCATGGCATATCTTAGATGTATCGGTTTATTACATGGCGAATAAAAAATATTATGCTTC 4694
Qy 2635 gtggaggaatttaagtgaactaatcgtaaatattccacttgggaatcagtgcgccaat 2694
Db 4695 GATTAGGGATATAAATTTATTCAACTATCGCTATGTACTTGGGAACCGGTGCGTCAAA 4754
Qy 2695 ccggtgtgaatgcagtaaaccaagacgcg---gggtagaactacacactcgatttggcgctc 2751
Db 4755 CAGCACAGGTGCGGTCAATCAACATCAAAATGTTGGTAGCTATATCTCGTACGACGAT 4814
Qy 2752 cggggagaaaatttcagtttagcatttgaatgaagtttta 2791
Db 4815 CAGGACGAAACTATACCTTAACATTAGAAATGAATTTCTA 4854
```

RESULT 10

```
V21430
ID V21430 standard; cDNA; 5009 BP.
AC V21430;
DT 20-JUL-1998 (first entry)
DE H. influenzae strain Minn A transferrin receptor operon.
KW tbp1; tbp2; vaccine; H. influenzae; antibody; diagnosis;
KW passive immunisation; ds.
OS Haemophilus influenzae.
FH key Location/Qualifiers
FT CDS 121..2103
FT /tag= a
FT /product= Tbp1
FT 2117..4855
FT /tag= b
FT /product= Tbp2
PN US5708149-A.
PD 13-JAN-1998.
PF 07-JUN-1995; 487890.
PR 08-NOV-1994; US-337483.
PR 08-NOV-1993; US-148968.
PR 29-DEC-1993; US-175116.
PR 07-JUN-1995; US-487890.
PR (CONN-) CONNAUGHT LAB LTD.
PI Chong P, Gray-Owen S, Harkness R, Klein M, Loosmore S,
PI Murdin A, Schryvers A, Yang Y;
PI WPI; 98-100410/09.
DR P-PSDB; W53046, W53847.
DR Purification of recombinant Haemophilus transferrin-binding protein
PT - by solubilising inclusion bodies separated from cell lysate
PS Example 4; Fig 5; 261pp; English.
CC The H. influenzae transferrin receptor operon contains two genes (tbp1
CC and tbp2) under the transcriptional regulation of one promoter. The
CC proteins encoded from these genes can be expressed in a recombinant
CC host. The proteins can be used in vaccines against H. influenzae
CC infections or to produce antibodies for use in diagnosis or passive
CC immunisation.
SQ Sequence 5009 BP; 1809 A; 836 C; 934 G; 1430 T;
```

Query Match 11.9%; Score 331.2; DB 1; Length 5009;
Best Local Similarity 50.5%; Pred. No. 4.7e-85;
Matches 1354; Conservative 0; Mismatches 1133; Indels 193; Gaps 15;

```
Qy 173 aaaaacgaagtaacttggtgggaaagtagtaaaaattatcacgaaatgagtaaaaaatc 232
Db 2307 ATATGAAGTAACCTGGACTGGCAAAATTTACAAACTAGTGAAGTATCAGCGCAGAAC 2366
Qy 233 aaattcttggtattcgtgatttaactcgctatgacctgtgatttcggtggtggaacaag 292
Db 2367 AAGTATTAAATATTCGTGATCTAACACGCTATGATCAGGGGATTTCAGTTGTAGACAAG 2426
```


Db	4455	TATTGATGCCATTCAGCCCGCGTTATATCATGTGGTTTAGGCTATGATCTCCAAGTA	4514
Qy	2465	aaaaatgggaattggcattactaccacacttctaaagccaaacccgcatgagtgg	2524
Db	4515	ATACTTGGGGAATTAATACAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGC	4574
Qy	2525	cagcacacgtcatcacgatacatgc-----gttgattaggtggcaaacctgcacg	2577
Db	4575	TAGGAAACGTGCTATGAGTGATGACAAATCAAGGGATGTAAATCAACAAGAAATCTTACT	4634
Qy	2578	gttttgggtacacccatgatattaccgg---ttacatcaattataaaactcacaccttac	2634
Db	4635	GGCATGGCATATCTTAGATGTATCGGTTATTACATGCGGAATAAAAAATATTATGCTTC	4694
Qy	2635	gtggaggaatttataatgtgactaactcgtaaatatccacitgggaatcagtcgcgcaat	2694
Db	4695	GATTAGGGATATATAATTTATTCAACTATCGGTATGTACTTTGGGAAGCGTGGCTGAAA	4754
Qy	2695	ccggtgtgaatgcagtaaacccaagacgg---gggtagcaattacacactcgattggcgctc	2751
Db	4755	CAGCACAAAGTGC GGCTCAATCAACATCAAAATGTTGTAGCTATACTGCTACGCAGCAT	4814
Qy	2752	cg99ggagaaatttcagtttagcatttgaattgaagtttta	2791
Db	4815	CAGACGAAACTATACCTTAACTATGAAATGAAATCTCTA	4854

RESULT 11

T49501 standard; DNA; 5033 BP.

AC T49501;

AD O5-MAY-1998 (first entry)

DE Transferrin receptor gene sequence encoding Tbp1 and Tbp2.

KW Transferrin receptor; Haemophilus influenzae type b;

KW Iron; human transferrin; iron source; antibody; bacterial growth;

KW vaccine; immunogenic truncated analogue; antigen; Tbp1; Tbp2; ss.

OS Haemophilus influenzae.

FH Key

FH Location/Qualifiers

FT CDS 169..2151

FT FT /tag= a

FT FT /note= "encodes Tbp2 (W08960)"

FT CDS 2129..4903

FT FT /tag= b

FT FT /note= "encodes Tbp1 (W08959)"

WO9640929-A2.

PN PD 19-DEC-1996.

PD 19-DEC-1996. CA0399.

PF 07-JUN-1996; PF 07-JUN-1996; US-649518.

PR 17-MAY-1996; US-483577.

PR 07-JUN-1995; US-483577.

PA (CONN-) CONNAUGHT LAB LTD.

PI Chong P, Gray-Owen S, Harkness RE, Klein MH, Loosmore SM,

PI Murdin AD, Schryvers AB, Yang Y;

DR WPI: 97-052329/05.

DR P-PSDB: W08959-60.

DR Haemophilus truncated transferrin receptor protein analogue, Tbp2 -

PT used to induce protection against disease caused by transferrin

PT producing pathogens, or as antigen to detect Haemophilus Tfr

PT antibodies

PS Claim 5: Fig 4A-Q: 228pp; English.

CC The present sequence represents the transferrin receptor gene of

CC Haemophilus influenzae type b, strain Eagan. The bacterial transferrin

CC receptor is composed of 2 chains, Tbp1 and Tbp2. H.influenzae is a

CC non-encapsulated or non-typable bacterium responsible for a wide range

CC of human diseases. Iron is an essential nutrient for the growth of these

CC bacteria, and they can utilise human transferrin as a source of iron.

CC Antibodies which block the access of the transferrin receptor to

CC its iron source prevent bacterial growth. The transferrin receptor, or

CC fragments, therefore, are good vaccine candidates. The full length Tbp2

CC protein is produced in low amounts in Escherichia coli. However, the

CC yield can be enhanced by truncation of the 3' end of the gene. An

CC immunogenic composition comprising (or encoding) the immunogenic

CC truncated analogue can be used to induce protection against a disease

Qy	1055	cgtctcattactcttggcgcaatctcttagaagatacaaaaaacgcgtaagatatccgtgata	1111
Db	3153	AACAACATTATATTTGGTGTGATTTTTTGAAATTCACACACAAAAATTTGATATCCGTGATA	3212
Qy	1115	tgaacgcgcacctactactatacaaaagacgatattaacttatcatctaggaaactagttt	1174
Db	3213	TGACATTTCCCGCTTATTTAAGCCCAACAGAAAGACGGGATGATAGTAGTCGTTCCTTTT	3272
Qy	1175	atgaagggataaatatttttagatggcttagttccaegccaaagatccccattggggtgc	1234
Db	3273	ATCCAATGCAAGATCATGTGCATATCAACATATTGAGGATGGCA-----GAGGCGTTA	3326
Qy	1235	gctatagccatgtgaagtttttttgatgaacgtcacccacaacacgcgtgtttaggattcaacct	1294
Db	3327	AATATGCAAGTGGCCTTTATTTTCGATGAACACCATAGAAAACACGCGTGTAGTATTGAAT	3386
Qy	1295	ataaatataaacacagagaataatcgctggttggtatagcattaaaactc---agtcggcata	1351
Db	3387	ATATTACGAAAAATAGAACAAGGGCGATCATTTGCAAAAGCAGTGTTAAGTGTCTAATC	3446
Qy	1352	aacaagatatgaactatatagcgggtcacatcgcttcattgctatgtagcattatcccctggtg	1411
Db	3447	AACAAMCATCATCTTGACACTATATCGGNACATACGCATTGCGCTCTTTATCCCTAATC	3506
Qy	1412	tagataaaattgcgccgcgactttggataaatcttgtctatgtatcgaactgaacgcgta	1471
Db	3507	CAAGTAAGAAATTGCCCCCAACACATTGATAAACCTTATTTCATATCTATCGTCTGATAGAA	3566
Qy	1472	ataattaccaagaagaagcatcgctgcattcattcatttagaattgataaa-----	1518
Db	3567	ATGTTTATAAGAAANAACATAATATGTTGCAATTGCAATTTAGAGAAAANAATTCACAAA	3626
Qy	1519	-gcgcctaagtgcgtcaagcgtatttaaaccacaccaaactgaattagggtgttg	1576
Db	3627	ATTGGCTTACTCATCAAAATGCTCTCAATCTCTGGTTTTGTATGACTTTACTCACGCGTTC	3686
Qy	1577	gctttgatcgatttaattcgcttaaggatcatggggtatgactgcccaatatataccaaag	1636
Db	3687	AGCATAAAGATTATTTAACTCGACGTGTTATCGCTACGCCAGATAGTATTCCAAGGAAC	3746
Qy	1637	gcggttataccagctaccgcggttagaggcgtttagataaaacctatatttatocgcgcg	1696
Db	3747	CTGGTGAAACTGTGTAACCAGAANAATGGTTTGCAATACAACTTACTTATPACCCAAAAC	3806
Qy	1697	atccagcagtatctgaacgcgtatcttctgttaataatacacgcgcgc-----gacatct	1750
Db	3807	CAGAGCCATNTTTCGAGACAGATCATGTTATTTATCAAGGTAGCTCTCTAATTACA	3866
Qy	1751	taaactgtgaacgcgctaaaaattaaaggcgatagccattttgttagcttcgcg----gac	1807
Db	3867	GAGACTGTAAAGTGCGGTTAATTTAAAGGGAAAAATTAATTTTCGACGACGCAATAATA	3926
Qy	1808	tagtgaataagcgsagtatgtgattttgggatttaggggtgcgttttttgatacaacatcgattta	1867
Db	3927	TGGCATTTAGGGAANAATACGTTGATTTAGGTTTAGTATTTCGGTATGACGTATCTCGTACAA	3986
Qy	1868	aattgatgatcgtgtgacacitaggccaacttatcgaaattgcttcgtgaatggtggga	1927
Db	3987	AAGTAATGAATCAACTATTAGTGTGGTAAATTTAAANAATTTCTCTTGGAAATTA	4046
Qy	1928	ttacgcttaaaccaacagaggtttgatcgctcttcttatcgcatattcaaacggttttagag	1987
Db	4047	TTGTCATAAAACCAACGGAATGCTTGATCTTCTTATCGCCTTCTACTGGAATTAGAA	4106
Qy	1988	tgcctgatctctatgaacttta tggtaaacgtgatacatattggggttaagaataacgaaat	2047
Db	4107	ATCCTAGTTTTTCTGAATATGATGTTGGTGGCGGTATGGTGGCAAGAANAATGACGAGGTTATG	4166
Qy	2048	atgtgaacgcgcgaacgtagccaccaggttagaccacgagaaaaatcgactaatcatgaga	2107
Db	4167	TAGTAAA-----TTTAGCCCTGAAMCATCTCGTAAACCAAGAGT	4205
Qy	2108	ttggagttagctttaaagggtcaatttgggttaccttgatgtgagctatttcogtataact	2167

Db	4206	TTGGTCTCGCTCTAAAGGGGATTTGGTAAATTAGATGAGATCAGTCATTTTAGTAAATGCTT	4265
Qy	2168	ataaaatgatgtgcgcagcagatgaaa--agaataatacaaaaatcacactgtttct	2224
Db	4266	ATCGAAATCTTATCGCTTGTCTGAAGAACTTAGTAAANAATGGAACTCGAAAGGCAATT	4325
Qy	2225	ataactaccataataatcaagatgtagcactaaaacgggataaaatttgcgtctaaattg	2284
Db	4326	ATGGATATCATAAATGCACAAAATGCAAAATTAGTTGGCGTAAATATAACTGCACAAATAG	4385
Qy	2285	acttacacggtatttcatctatcgtgcceagatgggttttatttcatcagttcgtctataacc	2344
Db	4386	ATTTTAATAGTGTATGGAACGATTTCCCTACGGTGGTATGCAACATTTGCTTATAAACC	4445
Qy	2345	gtgtaaaaagcaaaagcggaaactaacacgactcaagactcgatagcgtaaacagatccta	2404
Db	4446	AAGTAAAGTTAAGATC--AAAANAATCAATGCTGGTGTAGCCTCGGTAGCAGTATT	4502
Qy	2405	tcttagatgcgattcagccagcagcgcgtatgctgttgattcgggtacgatcacccagaag	2464
Db	4503	TATTTGATGCCATTCAGCCAGCCGTTATATCATTTGGTGTAGGCTATGATCATCCAACTA	4562
Qy	2465	aaaaatggggaattggcattactaccacctatctaaagccaaaacccgatgaggtgg	2524
Db	4563	ATACTTCGGGAATTAATACAATGTTTACTCAATCAAAAGCAAAATCTCAAATGAATTGC	4622
Qy	2525	caggcacacgtcatcgcgatactgcg-----gttgatttaggtggcaactgacgcg	2577
Db	4623	TAGGAAACGTGCTATTAGTAAACAATTAAGGGATGTAAATCACACAGAAAACTTACTC	4682
Qy	2578	gttcttggtcaccccatgatattaccgg---ttacatcaattataaaaactcacaccttac	2634
Db	4683	GGCATGGCATATCTTAGATGTATCGGTTATTACATGCGGAATAAAAAATATTATGCTTC	4742
Qy	2635	gtggaggaaatttaattgactgaactaactgtaaatattccacttgggaaacagtgcgcaat	2694
Db	4743	GATTAGGGATATATAATTTATCAACTATTCGCTATGTTACTTTGGGAACGGTGGCTCAAA	4802
Qy	2695	ccggtgtggaatcagtaaaacccaagacg--gggtagcaattacactcagttgttggcgtc	2751
Db	4803	CAGCAACAGTGGGGTCAATCAACATCAAAATGTTGGTACTATCTCGCTACGACGAT	4862
Qy	2752	cggggagaaatttcagtttagcatttgaattgaagtttta	2791
Db	4863	CAGACCAAACTATACCTTAACATTAAGAAATGAATTTCA	4902

RESULT 12

```

V21429
ID V21429 standard; cDNA; 5033 BP.
AC V21429;
DE 20-JUL-1998 (first entry)
DT H. influenzae strain Eagan transferrin receptor operon.
KW tbp1; tbp2; vaccine; H. influenzae; antibody; diagnosis;
KW passive immunisation; ds.
OS Haemophilus influenzae.
FH Key Location/Qualifiers
FT 169..2151
CDS
FT FT /*tag= a
FT FT /product= Tbp1
FT FT 2165..4903
FT FT /*tag= b
FT FT /product= Tbp2
US5708149-A.
PN PD 13-JAN-1998.
PF PD 07-JUN-1995; 487890.
PR PR 08-NOV-1994; US-337483.
PR PR 08-NOV-1993; US-148968.
PR PR 29-DEC-1993; US-175116.
PR PR 07-JUN-1995; US-467890.
PA (CONN-) CONNAUGHT LAB LTD.
PI Chong P, Gray-Owen S, Harkness R, Klein M, Loosmore S,

```

Pt Murdin A, Schryvers A, Yang Y;
DR WPI; 98-100410/09.
DR P-PSDB; W53044, W53045.
PT Purification of recombinant Haemophilus transferrin-binding protein
from sialylated inclusion bodies separated from cell lysate
PS Example 4; Fig 4; 261pp; English.
CC The H. influenzae transferrin receptor operon contains two genes (tbpl
and tbp2) under the transcriptional regulation of one promoter. The
CC proteins encoded from these genes can be expressed in a recombinant
CC host. The proteins can be used in vaccines against H. influenzae
CC infections or to produce antibodies for use in diagnosis or passive
CC immunisation.

SQ Sequence 5033 BP; 1822 A; 837 C; 940 G; 1434 T;

Query Match 11.9%; Score 331.2; DB 1; Length 5033;
Best Local Similarity 50.5%; Pred. No. 4.7e-85;
Matches 1354; Conservative 0; Mismatches 1133; Indels 193; Gaps 15;

Qy 173 aaacgaagtaactggttgggaaagttagtggaaaattatcacgaaatgagtaaaaaatc 232
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2355 ATAATGAAGTAACCTGGACTTGCGAAAATTATCAAAACTAGTGAAGATATCAGCGAGAAC 2414
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 233 aaattcttggtattcgatttaactgcgtatgacctggtatttcogtgggtgaacaac 292
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2415 AAGTATTAAATTCGTGATCTAACACGCTATGATCCAGGGATTTCAGTTGTGAACAAG 2474
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 293 gtcgsggtgcaagttagtggtgatgccatgcagggtgtagataaaaaaccggtgcagcttac 352
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2475 GTCGCGGTGCAGTCTCGATATCTATCTGCTGTGACAGAGAAATAGAGTTGCTTTAT 2534
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 353 ttgttgatgggtccaccaagcgacagtcattatcatacgtaggttc----- 398
Dy TAGTAGATGGTTTACCTCAAACGGAATCTTATGTAGTCAAAAGCCCTTTAGTGTGCTGTT 2594
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 399 ----agatgctaagtgtgtccaattaatgagattgagattgataaaaaattcgttcaattg 454
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2595 CAGGATATTCGCGACCTGGTGCAAATAATGAATGAATATGAATAATGAAGGCCGTCG 2654
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 455 agttaagcaaaaggagcaagttctgcgggaatatggctctggtgcgcatggtggtgctattg 514
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2655 AAATAAGCAAGGGGGGAGTCTCTGAGTATGGTAATGGAGACACTAGCTGGTCTGTAA 2714
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 515 gtlttcgtactaaagatgctgcaggatattttaagaaggggcagcaattggggcttagata 574
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2715 CATTTCAAGCAAAATCAGCAGCGCATATCTTTAGAAGGAGACAAATCATGCGGAATTCAAA 2774
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 575 gtaagacctttatgcagcaaaaaatagcattttttacgtctatcgcagcgcgtggtg 634
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2775 CTAAAAATGCTTTATTCAGCAAAAAATAAAGGCTTTACCCTTCTTTAGCTTAGCAGGAA 2834
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 635 agcgvggtggttttgaagcacttgttatgcaactcaccgcagcggtaaaagagaccacaaa 694
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2835 AACAAAGTGGATTTGAAGGGGTAGCCATTTACATCAACGAANAATCAATTGAAACCCAG 2894
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 695 ttcattccgagggcaataataataaacataatttcggcgctataccggcgtttgaaatc 754
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2895 TCCATAAAGATGCATTTAAAGCGGTACAAAGTATGATCGATTAATGCC----ACAACA 2950
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 755 gtcagcactttacccaaattccgcacagaatgctcctcgagagatctccttttaattggtg 814
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 2951 GATAAATCTTCAGGNATCTTTGTGATACAAGGTGAGTG-----TCCAATGGTG 2999
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 815 aagatacttgcaccaattagattgtactcctcgcgagggttaagtgaacgcgcgata 874
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 3000 ATGACAAGTGCAGC----- 3015
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 875 attccccagtgaacaatttcccgaatatatacgctgaagagcgcaaacagcttgagcaga 934
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dy 3016 -----CAAGCCACCTCGACTT 3032
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 935 ttccttatcgcaactgagcagctctcagcccaagatataccggttaaagatgcgattgcac 994
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```
QY 2048 atgtgcaacgcgcgaacgtagaccaccagtttagagccagcaaaaaatcgactaatcatgaga 2107
Db 4167 TAGGTAATA-----TTTAAAGCCTGAAACATCTCTGTAACCAAGAGT 4205
QY 2108 ttggagttagctttaaaggtcaatttggttacctttagtgagactatccgtaataaact 2167
Db 4206 TTGGTCTCGCTCTAANAAGGGGATTTTGGTAATATTCAGATCAGTCATTTTAGTAATGCTT 4265
QY 2168 ataaaaatgattgagcagcatgtataa---agaataatacaaaaaatcacactgtttct 2224
Db 4266 ATCGAAATCTTATCGCCTTTGCTGGAAGAACTTAGTAATAAATGAACTGGAAGGGCAATT 4325
QY 2225 ataaactacataataattcaagatgtagcactaaacgggataaatttagtcgtataatttg 2284
Db 4326 ATGGATATCATATGCAAAATGCAAAATAGTTAGTGGCGTAAATATTAACGTACCAATTAG 4385
QY 2285 acttacacaggtatttatcatgtccagatggtttttattcatcagttgctataaac 2344
Db 4386 ATTTTAAATGGTTATGGAACGTTATCCCTACGGTTGGTATGCAACATTTGCTTATAACC 4445
QY 2345 gtgtaaaagtataaagcggaaactaacccgactcaagactcgatagcgtataaacctcta 2404
Db 4446 AAGTAAAGTTAAAGATC---AAAAATCAATGCTGTTTAGCTCCGTAAGCAGTTATT 4502
QY 2405 ttctagatgcgattcagccagcagcgtatgtctgttgattcggctacgatcacccagaag 2464
Db 4503 TATTGATGCCATTCAGCCGAGCCGTTATATCATCTGTTAGGCTATGATCATCCAAAGTA 4562
QY 2465 aaaaatgggaattggtcattactaccacactattctaaagccaaaacgcgcatgaggtgg 2524
Db 4563 ATACTTTGGGAATTATACAAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGC 4622
QY 2525 caggcacacatcatcgcgatacatgc-----gttgatttagtggcaaaactgaccg 2577
Db 4623 TAGGAAACGCTGATTAGGTAAACATTAAGGGATGTAATAACAAAGAAACTTACTC 4682
QY 2578 gttcttggtacacctgatattaccgg---ttacataaattataaaaactacacaccttac 2634
Db 4683 GGGCATGGCATATCTTAGATGTATCGGGTTATTACATGGCGAATAAAATATTATGCTTC 4742
QY 2635 gtggaggaatttataatgtgactaatcgtataatctccacttggtggaatcagtcgcgaat 2694
Db 4743 GATTAGGATATATATTATTCAACTATCGCTATGTTACTTTGGGAAGCGGTGCGTCAA 4802
QY 2695 cogggtgtaatgcagtaaaccaaacgcg---gggtagcaattacacactcgatttggcgctc 2751
Db 4803 CAGCACAAAGTGGCGGTCAATCAACATCAAAAATGTTGGTAGCTATACTCGCTAGCGCAT 4862
QY 2752 cggggagaaatttcagtttagcatttgaatgaagttaa 2791
Db 4863 CAGGAGAAACTATACCTTAACATTAGAATGAATTTCTA 4902

RESULT 13
Q94446
ID Q94446 standard; DNA; 5144 BP.
AC Q94446;
DE Bacterial transferin receptor operon (H. influenzae strain SB33).
KW Tbp1; Tbp2; transferin receptor operon; vaccine; antigen;
KW non-typable strain; Haemophilus influenzae; meningitis; ss.
OS Haemophilus influenzae strain SB33.
FH Key
FT Location/Qualifiers
FT cds
FT /*tag= a
FT /product= Tbp2
FT 2135..4870
FT /*tag= b
FT /product= Tbp1
FT
FT W09513370-A1.
PD 18-MAY-1995.
PF 07-NOV-1994; CA0616.
```

```
PR 08-NOV-1993; US-148968.
PR 29-DEC-1993; US-175116.
PA (CONN-) CONNAUGHT LAB LTD.
PI Chong P, Gray-owen S, Harkness R, Klein M, Loosmore S;
PI Murdin A, Schryvers A, Yang Y;
DR WPI: 95-194089/25.
DR P-PSDB: R77891-92.
DR Nucleic acids encoding Haemophilus transferrin receptor - used to
PT develop prods for detection and in diagnosis, prevention and
PT treatment of Haemophilus infection.
PS Claim 12; Fig 7A-N; 231pp; English.
CC The present DNA shows the transferin receptor (Tfr) operon consisting of
CC two genes (Tbp1 and Tbp2) arranged in tandem and which are transcribed
CC from a single promoter, from the non-typable Haemophilus influenzae
CC strain SB33. The SB33 Tbp2 gene has a single base deletion of the resulting
CC in a frame-shift at residue 126 and premature truncation of the resulting
CC protein at residue 168. H. influenzae Tfr is iron- and/or haemin-
CC regulated and a putative fur-binding site has been identified upstream of
CC Tbp2. Antibodies blocking this binding site may prevent bacterial growth.
CC Fragments of the Tfr (or its genes) are useful in vaccines to provide
CC protection against, e.g. bacterial meningitis. An advantage of using the
CC Tfr is that it shares homology with Tfr of other H. influenzae strains
CC including non-typable strains.
SQ Sequence 5144 BP; 1813 A; 888 C; 973 G; 1470 T;

Query Match 11.9%; Score 331; DB 1; Length 5144;
Best Local Similarity 50.5%; Pred. No. 5.4e-85;
Matches 1352; Conservative 0; Mismatches 1135; Indels 190; Gaps 15;

QY 173 aaaaagaactgaactggcttggggaagtagtaaaaaattatcacgaataagtagtaaaatc 232
Db 2325 ATAATGAAGTAAGTGGACTTGGCAAAATTTATAAAACGAGTGAAGATATCAGCCGAGAAC 2384
QY 233 aaattcttggtattcgtattaaactcgctatgacccctggtatttcgggtggtagaacaag 232
Db 2385 AAGTATTAATTCGTGATCTAACACGCTATGATCCAGGCATTTTCAGTTGTAGAACAA 2444
QY 293 gtccgggtgcaagtagtggtatgcattcgaggtgtagataaaaaacccgtgtagcttac 352
Db 2445 GTCCGGTGCAGATTCGTGATATTCATTCTGTTGTTAGTGGACAGAAATAGAGTTGCTTAT 2504
QY 353 ttggtgatgggtaccacagcagcagattatcatcagctagttc----- 398
Db 2505 TAGATAGTGGTTACCTCAACGCAATCTTATGTAGTGAAGCCCTTTAGTTGCTGCTT 2564
QY 399 ----agatgctaagtgggtgcaattaatgagattgagtagtaaaacattcgctcaattg 454
Db 2565 CAGGATATTCTGCACTGCTGCAATTTAATGAAATTCATGAAATGTAAAGCCGCTG 2624
QY 455 agttaagcaaaagagcaagttctgctggaatatgctcctgggtgcatggtggtgctattg 514
Db 2625 AAATAAGCAAGGGGGGAGTTCTTCTGAGTATGTTAATGAGCAGTGTGTTGTTCTGTAA 2684
QY 515 gtttcgtactaaagatgcgaggtattataaagaggggcagcattggggttagata 574
Db 2685 CATTTCAAGCAAAATCCGCGCGATATCTTAGAAGAGACAAATCATGGGAATTCAAA 2744
QY 575 gtaagacctcttatgccagcaaaaaatagccattttttacagttatcgagcggtgggtg 634
Db 2745 CTAAAAATGCTTATTCAAGCAAAATAAAGGCTTTTACCATTCTTTAGCTGTAGCAGGAA 2804
QY 635 aggcgggtggttttgaagcaacttcttattgcaactcacccagcagcgtataagagacaaaa 694
Db 2805 AACAGGTGATTTGAAGGGGTGCGCATTTTACACTCAAGCAAAATTCGAGGAAACCAAG 2864
QY 695 ttctccagagcaataataataataatttcggtatattcggtatataccggctttagaatc 754
Db 2865 TCCATTAAGATGCAATTAAGGCGGTACAAAGTTATGAGCGATTCATCGCC-----ACAACA 2920
QY 755 gctacgactttaccacaaattccgcagacagatgctcctggaggaggtatcctcttttaattg 814
Db 2921 GATAAATCTCAGGATACCTTTGTGATACAAAGGTGAGTGTG-----TCCAATGGTG 2969
```

Qy 815 aagatacttgcccaacattagattgactcctcgtcgaaggttaagttgaaccgcgata 874
Dy 2970 ATGACAAGTGTGCAGC-----2985
Qy 875 atttccagtgagaacatttcgcggaatatacgcctcgaagagcgcaaacagagttgagcaga 934
Dy 2986 -----CAACCCACTGCAAAAGT 3002
Qy 935 ttccttatcgcactgagcagctctcagcccaagaataataccggtaaaagatcgcattgcac 994
Dy 3003 TATCCCCCAAGCGAAGCCGTAACGGTTTACAGATTATACGGGGCTAACCGTATCAAC 3062
Qy 995 caaaccttttagattacaagagtaattctgtttttatgaagtttgctatcaactcaact 1054
Dy 3063 CTAATCCCAATGAATATGAAGCCAGCTGTGGTTTAAAGAGGGGTATCATTTTCTG 3122
Qy 1055 cgtctcattatctcgcgaacttttagagatacaaaaaacgcgtacgatatacgtgata 1114
Dy 3123 AACAACTATATTGGTGGTATTTTGAATTCACACAAACAAAATTTGATATCCGTGATA 3182
Qy 1115 tgcgaacccagcttactatacaaaagacgatatataacttatcacttaggaactatgttt 1174
Dy 3183 TGACATTTCCCGCTTATTAGATCAACAGAAAACGGGATGATAGAACTGGCCCTTTT 3242
Qy 1175 atgaaggggataaatttttagatggcttagtggttcaagccaaggtaccccttatgggtgc 1234
Dy 3243 ATCCAAAGCAAGATTATGGTGCATATCAACGTATTGAGGATGGCCGA-----GSCGTTA 3296
Qy 1235 gctatagccatgtgaagttttttgatgaagcgtcaccacaacacgctgtttagattcaact 1294
Dy 3297 ATATGCAAGTGGGGCTTTATTTCCGATGAACACCAATAGAAAACAGCGTGTAGTATTGAAT 3356
Qy 1295 ataaataaaacagagaataatcgtggttggatagcattaaac---tcagtcgggata 1351
Dy 3357 ATATTACGAAAATPAAAGCAAAACGCGGCATCAITGACAAAGCAGTGTAAAGTGTAAATC 3416
Qy 1352 aacaagatttgaactatagaccgggtcacatcgttcattgttagcagattatcctctgg 1411
Dy 3417 AACAAACATCATCTTGACAGTTATATGCGACATACGCAATTGCGAGTCTTTATCCTAATC 3476
Qy 1412 tagataaaattcgcgcgcactttgggataaaattcgttctatgtatogaaactagagcta 1471
Dy 3477 CAAGTAAGAAATGCGCCGACACTGTGATAACCTTATTCATCTACTCTGTTCTGATAGAA 3536
Qy 1472 ataatccaagaagaacatcgtgctcattcatttagaatttgataaagcgtaaatgctg 1531
Dy 3537 ATGTTTATAAGAAAACATATATGTTGCAATTGAAATTTAGAGAAAATAATCAACAAA 3596
Qy 1532 gtcaagggcgtatttaaccaaaccacaactgaatttaggg-----ttgggctttg 1582
Dy 3597 ATGGCTTACTCATCAAAATTGCTTCAATCTGGTTTGTGAGACTTTTACTTCAGCGCTC 3656
Qy 1583 atcgatttaattcgtttagatcgaatgggatactgactgcccataataccaaaggcggtt 1642
Dy 3657 AGCATAAAGATTATTAATCTCGACGTGTACCGCTACGGCAAAATATTATTTTCAGGGACAG 3716
Qy 1643 ataccagctacccggtagaggggttttagataatcaatatatttatcgcgcgatccac 1702
Dy 3717 TTGCTGGTAAACGAAGAATGGTTACGAAAACAACTTACTTATCTACFCAAAACCAAAAG 3776
Qy 1703 gcagatttgaacggtatctctgttgaataatacacgcgcg-----gacatcttaaaact 1756
Dy 3777 TAGATTTTGTAGGACAAATCATTTGATATATTAAGGTAGTCTCTCTAAATTACAGCGACT 3836
Qy 1757 gtgaacgcgtgaaatttaaaagcgatagccattttgttagcttcgc---gatctagtga 1813
Dy 3837 GTAAAGTCGGTTAATTAAGGGAAAAATTTATTTTCGCAGACGCAATAATATATGGCAT 3896
Qy 1814 taagcgaatgttgatttgattagggggttgcttttgatcaacatcgatttaaatctg 1873
Dy 3897 TAGGAAATACATGATTAGGTTTAGGTATTCGGTATGACGTATCTCGTACAAAAGCTA 3956

Qy 1874 atgatccgtggacacattagccgaacttatcgaaaatttgcttggtaagtgtggattacgc 1933
Dy 3957 ATGAATCAACATATTAGTGTGGTAAATTTAAATAATTTCTCTTGGAAATACTGGTATTGCA 4016
Qy 1934 ttaaaccaacagagatttgatcgcgttttcttatcgcattcaaacggtttttagagtgccgtg 1993
Dy 4017 TAAACCAACAGGAATGGCTTGATCTTTCTTATCGCTTTCTTACTGGATTATGAATCCTA 4076
Qy 1994 cattctatgaactttatggttaaacgctgatcatatttgggcttaaaagataaacgaatatgac 2053
Dy 4077 GTTTTGTGTAATGTATGGTTGGCGGTATGGTGCATAATATAGCGATGTTTATGTAGTA 4136
Qy 2054 aacgcgcgaacgttagccaccagtttagagccagaaaaatcgactaatcatgatgattggag 2113
Dy 4137 AA-----TTAAGCCTGAAACATCTCGTAACCAAGAGTTGGTC 4175
Qy 2114 ttacgtttaaagggtcaatttggttaccttgatgtgagcgtatttccgttaataaactataaaa 2173
Dy 4176 TCGTCTAAAGAGGGGATTTTGGTAAATATTGAGATCAGTCATTTTAGTAATGCTTATCGAA 4235
Qy 2174 atatga-----ttgcgcacagcatgtaaaagaataatacaaaaaatcacactgtttctata 2227
Dy 4236 ATCTTATCGCCTTTGCTGACGAACCTTAGTAAAAATGGAACACTCTGGAAGGGCAATTATG 4295
Qy 2228 actaccataattcaagatgtagcactaaacgggataaaattttagtcgctaaaatttgact 2287
Dy 4296 GATATCAATATGACAAAATGCAAAATTAGTTGGCGTAAATATATAACTGCGCAATTAGATT 4355
Qy 2288 tacacgtattttatctatcgtcgcagatgggttttttattcatcagttgcttataaacogtg 2347
Dy 4356 TTAATGTTTTATGAAAACGTATTCCCTACGTTGGTATGCAACATTTGCTTTATAACCGAG 4415
Qy 2348 taaaagtaaaagagcgaaactaaaccgacacgaacgcgtatagcgtataaagatcctcttc 2407
Dy 4416 TAAAAGTTAAAGATC---AAAAATCAATGCTGGTTGGCCTCCGTAGCAGTATTATTAT 4472
Qy 2408 tagatgcgattcagccagcacgctatgtcttgattcgcgctacgcatcacccagaagaaa 2467
Dy 4473 TTGATGCCATTTACGCCAGCGCTTATATCATTTGTTTAGGCTATGATCATCCAAGTAATA 4532
Qy 2468 aatggggaattggcattactaccacctattttaaagccaaaaacgcgcgatgagtggcag 2527
Dy 4533 CTTGGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGCTAG 4592
Qy 2528 gcacacgtcatccaggtatcacatgc-----gttgatttaggtggcaaacctgaccggtt 2580
Dy 4593 GACAACGTGCTATTGGGTAAACAATTCAGGAATGTAAATCAACAGAAACTTACTCGGG 4652
Qy 2581 ctgtgtacacccatgatattaccgg-----ttacatcaattataaaactcacaccttagctg 2637
Dy 4653 CATGCGATATCTTAGATGTATCGGGTTATTACATGGCGAATAAAAATATTATGCTTCGAT 4712
Qy 2638 gaggaaattataatgtgactaaatcgttaaatattccattcgttggaatcagtgcccaatccg 2697
Dy 4713 TAGGGATATAATTAATTAACATATCGCTATGCTTACTTGGGAAGCGGTGCGTCAACAG 4772
Qy 2698 gtgtgaatgcagtaaaacaaagacc---ggggtagcaattacactcgtatttggcgctcgg 2754
Dy 4773 CACAAGTGGCGGTCAATCAACATCAAAATGTTGTAGCTATACTCTCGTACGACGATCAG 4832
Qy 2755 ggagaaatttcagtttagcatttgaagtgtta 2791
Dy 4833 GACGAACATATACCTTAACATTAGAAATGAATTTCTA 4869

RESULT 14

Q94444

ID Q94444 standard; DNA; 5009 BP.

AC Q94444:

DT 09-OCT-1996 (first entry)

DE Bacterial transferrin receptor operon (H. influenzae strain Minna).

KW Tbp1; Tbp2; transferrin receptor operon; vaccine; antigen;

KW non-typable strain; Haemophilus influenzae; meningitis; ss.


```
Qy 1751 taaactgtgaacccgtaaaattaaagcgtagccattttagcttcgcg---gatc 1807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3819 GAGACTGTAAAGTGGGTTAATAAAGGAAAAATTAATTTTCGACGACGCAATAATA 3878

Qy 1808 tagtgataagcagatgattgattgggattaggggtcggtttgatcaaatgattta 1867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3879 TGGCATTAGGGAATAACGTTGATTAGTTAGTTAGTTATTCGTTATGACGATCTCGTACAA 3938

Qy 1868 aatcgtgatccggtggacactgacgaacttatcgaaatgtgcttgggaatggtgga 1927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3939 AAGCTAATGAATCAACTATTAGTGTGTAATTAATAAATTCCTTGGAACTACTGCTA 3998

Qy 1928 ttacgcttaaaccaacagattgttatcgcttcttcttatcgacttcaaacggttttagag 1987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3999 TTGTCATAAAACCAACGGAATGGCTGATCTTCTTATCGCTTCTTACTGGAATTAGAA 4058

Qy 1988 tgcgtcattctatgaactttatggttaaacgtgatcatattgggcttaaaagataacgaat 2047
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4059 ATCCTAGTGTCTTCTGAAATGATGTTGGCTGGCGTATGGTGGCAAGAAATGACGAGGTTTATG 4118

Qy 2048 atgtgcaacgcgcgaacgttagccaccagttagagccagagaaaatcgactaatcataga 2107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4119 TAGGTAAA-----TTTAAAGCTGAAACATCTCGTAACCAAGAGT 4157

Qy 2108 ttggagtttagcttaaaagttcaattggttaccttggatgtgagctatttccgtaataaact 2167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4158 TTGGTCTCGCTCTAAAGCGGATTTGCTAATATTGAGATCAGTCATTTTAGTAATGCTT 4217

Qy 2168 ataaaaatatgattgacagacatgtaaa---agaataatacaaaaaatcaactgtttct 2224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4218 ATCGAAATCTTATCGCCCTTGTCTGAAGAACTTAGTAAAAATGGAATCGAAAGGCAATT 4277

Qy 2225 ataaactacaataattcaaatgtagcactaaacgggataaatttagcttaaatgtg 2284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4278 ATGGATATCAATAATGCACAAATGCAAAATAGTTGGCGTAAATATAACTGCACAAATTAG 4337

Qy 2285 acttacacggttttttctatgctgcagatggttttttattcattcagttcttataacc 2344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4338 ATTTTAATGGTTTATGGAAAGCTATTCCTACGGTGTGTGATGCAACATTTGCTTATAACC 4397

Qy 2345 gtgtaaagttaaaagcggaaactaacccgactcaagactcgatagcgtaaacatcccta 2404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4398 AAGTAAAAAGTTAAAGATC---AAAAATCAATGTGTTAGCCTCCCGTAAGCAGTATT 4454

Qy 2405 ttctagatgcattcagcagcagcagctatgcttggtgattcggtacatcacccagaag 2464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4455 TATTGTATGCCATTACGCCCGCCGCTTATCATTTGTTAGGCTATGATCATCCCAAGTA 4514

Qy 2465 aaaaatggggaattggcattactaccactattctaaagccaaacacgccgatggtgg 2524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4515 ATACTTGGGAATTAATCAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGC 4574

Qy 2525 caggcaacgcgtcaatcagcgatacatcgc-----gttgattaggtggcaaacgacgcg 2577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4575 TAGGAAAACGTCATTAGGTAACTAATCAAGGGATGTAATAATCAACAAGAAAATTTACTC 4634

Qy 2578 gtcttgatcacccatgatattaccgg---ttaacatcaattataaaactacaccttac 2634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4635 GGGCATGGCATCTTATGATGATCGGGTTTATACATGGCGAATAAAATATTATGCTTTC 4694

Qy 2635 gtggagggaatttataatgtgactaatcgtaaatattccacttgggaatcagtgccccaat 2694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4695 GATTAGGGATATATAATTATTCACTATCGCTATGTTACTTGGGAAGCGGTGCGTCAAA 4754

Qy 2695 ccggtgtgaatgcagtaaaccaagaccg---gggtagcaattacactcgatttggcgctc 2751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4755 CAGCAACAGGTGCGGTCAATCAACATCAAAATGTTGGTAGCTATATCTCGCTACGCAGCAT 4814

Qy 2752 cggggagaaaattcagtttagcattgtaaatgaagtttta 2791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4815 CAGGAGGAAACTATACCTTAAACATTAGAAATGAAATTTCTA 4854
```

RESULT 15

```
Q94443
ID Q94443 standard; DNA; 5033 BP.
AC Q94443
DT 09-OCT-1996 (first entry)
DE Bacterial transferrin receptor operon (H. influenzae strain Eagan).
KW Tbp1; Tbp2; transferrin receptor operon; vaccine; antigen;
KW non-typable strain; Haemophilus influenzae; meningitis; ss.
OS Haemophilus influenzae strain Eagan.
FH Key Location/Qualifiers
   cds 169..2151
      /tag= a
      /product= Tbp2
      /tag= b
      /product= Tbp1
   PN W09513370-A1.
   PD 18-MAY-1995.
   PF 07-NOV-1994; CA0616.
   PR 08-NOV-1993; US-148968.
   PR 29-DEC-1993; US-175116.
   PA (CONN-) CONNAUGHT LAB LTD.
   PI Chong P, Gray-owen S, Harkness R, Klein M, Loosmore S;
   PI Murdin A, Schryvers A, Yang Y;
   PI P-PSDB: R77885-86.
   DR Nucleic acids encoding Haemophilus transferrin receptor - used to
   PT develop prods for detection and in diagnosis, prevention and
   PS treatment of Haemophilus infection.
   PS Claim 12; Fig 4A-Q; 231pp; English.
   CC The present DNA shows the transferrin receptor (TfR) operon consisting of
   CC two genes (Tbp1 and Tbp2) arranged in tandem and which are transcribed
   CC from a single promoter, from Haemophilus influenzae type b, strain Eagan.
   CC H. influenzae TfR is iron- and/or haemin-regulated and a putative fur-
   CC binding site has been identified upstream of Tbp2. Antibodies blocking
   CC this binding site may prevent bacterial growth. Fragments of the TfR
   CC (or its genes) are useful in vaccines to provide protection against, e.g.
   CC bacterial meningitis. An advantage of using the TfR is that it shares
   CC homology with TfR of other H. influenzae strains including non-typable
   CC strains.
   SQ Sequence 5033 BP; 1822 A; 840 C; 937 G; 1434 T;

Query Match 11.8%; Score 329.6; DB 1; Length 5033;
Best Local Similarity 50.5%; Pred. No. 1.4e-84;
Matches 135; Conservative 0; Mismatches 1134; Indels 193; Gaps 15;

Qy 173 aaacgaagtaactggcttgggaaagtagtgaataattatcacgaaatgagtaaaatc 232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2355 ATATGAAGTAACCTGGACTTGGCAAAATATCAAACTAGTGAAGTATCAGCCGAGNAC 2414

Qy 233 aaattcttggtattcgtgatttaactcgtatgaccttggtatttcggtggtgaaacaag 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2415 AAGTATTAAATATTTCGTGATCTAACACGCTATGATCCAGGGATTTTCAGTTGTAGAACAAG 2474

Qy 293 gtccggtgcaagtagtggctatgccattccattcaggttagataaaacccgtgcagcttac 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2475 GTCCGGTGCANGTTCGGATATTCTATTCTGTTATGGACAGAAATAGATTGCTTTAT 2534

Qy 353 ttgttgatgggctaccacaagcgacaggttatcatcacgtaggttc----- 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2535 TAGTAGATGGTTTACCTCAACGCAATCTTATGTAGTGAAGCCCTTTAGTTGCTCGTT 2594

Qy 399 ----agatgctaatggtggtgcaattaatgagattgagattgaaacattggttcaattg 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2595 CAGGATATTCTGGCACCTGGTGCATTAATGAATTAATGAATGAATAATGAAGGCGGTCG 2654

Qy 455 agttaagcaaaaggaagattctcgcgaataatgctctggtgcgcagtggtggtcatttg 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2655 AAATAAGCAAGGGGGGAGTTCCTTCTGATGATGTAATGGAGACTAGCTGTTCTGTAA 2714

Qy 515 gtttcgtactaaagatgcgcaggtattattataaaggggcagcattggggttagata 574
```

Db 2715 CATTTCAAAGCAAAATCAGCAGCCGATATCTTTAGAAGGAGACAAATCATGGGGAATTCAAA 2774
Qy 575 gtaagaccttattgccaagcaaaaatagccatttttttacagctctatcgagcggtggtg 634
Db 2775 CTAATAATGCTTATTTCAAGCAAAAATAAGGCTTTTACCATTCTTTAGCTGTAGCAGAA 2834
Qy 635 agcggtggttttgaagcactgttttgaactccgacacggttaagagaccacaaa 694
Db 2835 AACAAAGTGGGATTTGAAGGGCTAGCCATTTACACTCAACGAANAATCAATTTGAAACCCAAAG 2894
Qy 695 ttcatctcgagggcaaaataaataaaacataattcggcggtataaccggctttgaaatc 754
Db 2895 TCCATAAAGATGCATTAAGGCGGTACAAAGCTTATGATCGATTAAATCGCC-----ACAAACA 2950
Qy 755 gctacgacctttaccocaaattccgcacagaaatgctccctggaggatctcctttaaattg 814
Db 2951 GATAAATCTTCAGGATACTTTGTGATACAAGGTGAGTG-----TCCAAATGGTG 2999
Qy 815 aagatacttgcccaacattagattgtactcctcgtgcaagggttaagttagaacgcgata 874
Db 3000 ATGCAAGTGTGCAGC----- 3015
Qy 875 atttcccgatgagaacatttccggaatatatagcctgaagagcgcaaacagcttgagcaga 934
Db 3016 -----CAAGCCACTCGGACTT 3032
Qy 935 ttcctattcgcactgagcagctctcagcccaagaataatcaccggttaagatgcattgcac 994
Db 3033 TATCCACCAAGGAAACCCGTAAGCGCTTCAGATTATACGGGGCTAACGGTATCAAAAC 3092
Qy 995 caaacctttagattacaagagtaattcgtttttatgaagtttggctatacactcaact 1054
Db 3093 CTAATCCAAATGAATATGAAGCGCAGCTCTGGTTTTTAAAGAGGAGGTATCATTTTTCTG 3152
Qy 1055 cgtctcattacttggcgcaactcttagagatacaaaaacgcgtacgatatccgtgata 1114
Db 3153 AACAACTATTATTTGGTGTATTTTGAATTCACACACAAANAATTTGATATCCGTGATA 3212
Qy 1115 tgcacaacgcgcctactatatacaaaagacgataattaacttataccttaggaactatgtt 1174
Db 3213 TGACATTTCCCGCTTATTTAAGCCCAACAGAAAGACGGGATGATAGTAGTGGTCTCTTTT 3272
Qy 1175 atgaagggtataatttttagatggcttagtggttcagacccaagatcccttatgggtgc 1234
Db 3273 ATCCAAATGCAGATCATGGTGATATCAACATATTGAGGATGGCA-----GAGGCGGTTA 3326
Qy 1235 gctatagccatgtgaagtttttttgaagcgtcaccaacacgcgtctttaggattcacct 1294
Db 3327 AATATGCAAGTGGGCTTTATTTTCGATGAACACCATAGNAACACGCGTGTAGGTATTGAAT 3386
Qy 1295 ataaataaaaccagagaataatcgtggttggatagcattaaactc---agtcgcgata 1351
Db 3387 ATATTTAGAAAATAAGAACAAAGCGGCATCATTTGACAAAAGCAGTGTAAAGTGTCTAATC 3446
Qy 1352 aacaagatttgaactatataagcggctacacgtctgcatgttagcattagcattatcctgtg 1411
Db 3447 AACAAACATCATCTTGACAGTTATATGCGACATACCGCATTTGCGATCTTTATCCTAATC 3506
Qy 1412 tagataaaaattgcgcgcgcactttgggataaattcttggctctatgtatgaactgagcgt 1471
Db 3507 CAAGTAAGAATTCGGCCCAACACTTGTATAAACCTTATTCATCTATCGTTCTGTAGAA 3566
Qy 1472 ataattaccagaagaagcatcgtgtcattcattttagaatttgataaa----- 1518
Db 3567 ATGTTTAAAGAAAACATAATATGTTGCAATTTGAAATTTAGAGAAAAAATAATTCACAAA 3626
Qy 1519 --gcgctaaatgctgggtcaagcgctatttaaccacaaacccacaaactgaatttaggttg 1576
Db 3627 ATTGGCTTACTCATCAAAATGTTCTTCAATCTTGTTTGTATGACTTTACTTCAGCGCTTC 3686
Qy 1577 gctttgatcgtatttaattcgtcttatggatcatggggatgactgcccataataccacaa 1636

Db 3687 AGCATAAAGATTATTTAACTCGAGCTGTTATCGCTACGCGCAGATAGTATTCCAAGGAAAC 3746
Qy 1637 gcggttataccagctaccgcggttagagggcgttttagataataatccatatattatcgccgcg 1696
Db 3747 CTGGTGAACCTGCTAAACCAAGAAATGTTTGAATCACAACCTTACTTATACCAAAAC 3806
Qy 1697 atccagcagatttgaacgggtatcttcttgtaataacacacgcggc-----gacatct 1750
Db 3807 CAGAGCCATATTTTGCAGGACAGATCATTTGTAATATCAAGGTAGCTCTCTTAATTACA 3866
Qy 1751 taaactgtaacgcgcgtataaattaaagcgatagccattttgttagcttccgc---gac 1807
Db 3867 GAGACTTAAAGTGCAGTTAATTAAGGGAATAATTAATTTTCGCAGCAGCAATAATA 3926
Qy 1808 tagtgaataagcaggtatgctgatttgggttaggggtgcgttttttgatacaacatcgattta 1867
Db 3927 TGGCATTTAGGGAATAACGTTGATTTAGTGTAGGTATTCGGTATGAGTATCTCGTACAA 3986
Qy 1868 aatctgatgatccgtggacacttagccgaactatcgaaattggcttggaaatggtggga 1927
Db 3987 AAGCTAATGAATCAACTATTAGTGTGGTAAATTTAAAAATTTCTCTTGGAAATACTGGTA 4046
Qy 1928 ttagccttaaaccaacagagtttggatcgcctttcttatcgcatttcaaacgggttttagag 1987
Db 4047 TTGTCATAAAACCAACGGAATGCTTGATCTTCTTATCGCCTTTCTACTGGATTTAGAA 4106
Qy 1988 tgcctgactctatgaactttatggttaaacgctgatactat1tggccttaagagataaacgaat 2047
Db 4107 ATCCTAGTTTTTCTGAAATGATGTTGGGGTATGTTGGCAAGAAATGACGAGGTTTATG 4166
Qy 2048 atgtgcacgcggcgaacagtagccaccagttagccagacagaaaaatcgactaatcataga 2107
Db 4167 TAGTAAA-----TTAAGCCTGAAACATCTCGTAACCAAGAGT 4205
Qy 2108 ttggagttagctttaaagggtcaatttggttacccttgatgagctatttccgttaataact 2167
Db 4206 TTGGTCTCGCTCTAAAGGGGATTTTGGTAATATTGAGATCAGTCAATTTTAGTAATGCTT 4265
Qy 2168 ataaaaata1gattgcgacagcagcatgtaaa---agaaataacaaaaatccactgtttct 2224
Db 4266 ATCGAAATCTTATCGCCTTTGCTGAAGAACTTAGTAAAAATTGGAAATGGAAGGCAATT 4325
Qy 2225 ataactaccataatactcaagatgtagcactaaacgggataaaatttagtgcgtctaaattg 2284
Db 4326 ATGGATATCATATATGCAAAATGCAAAATAGTTGGCGGTAAATATTAAGTCACAAATAG 4385
Qy 2285 acttacacggtattttatctatgctgccagatggttttttattcatcagttgcttataacc 2344
Db 4386 ATTTTAATGGTTTATGGAACGATATTCCTACGGTTGGTATGCAACATTTGCTTATRAAC 4445
Qy 2345 gtgtaaaagtataagagcggaacactaacgcactcaagactcgatagcgtaaacgatccta 2404
Db 4446 AAGTAAAAGTTAAAGATC---AAAAATCAATGCTGTTTAGCCTCGTAAAGCAGTATT 4502
Qy 2405 ttctagatgctgattcagcgacgctatgcttgcattcggtcagctagcaccacagaag 2464
Db 4503 TATTTGATCCCATCTACCCAGCGCTTATATCATTTGTTTAGGCTATGATCATCCAGTA 4562
Qy 2465 aaaaatgggaaattggtcattactaccacctatttctaagccaaaacgcgcgatgagtg 2524
Db 4563 ATACTTGGGGAATTAATACAATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTGC 4622
Qy 2525 caggcacagctcatcagggatacatcgc-----gttgatttaggtggcgaactgaccg 2577
Db 4623 TAGGAAACCGTGCATTAGGTAACAATTTCAAGGGATGTAAAAATCAACAGAAATTTACTC 4682
Qy 2578 gttcttggtcacaccatgatattaccgg---ttacatcaaatataaaaaactacacaccttac 2634
Db 4683 GGGCATGGCATATCTTAGATGTATCGGTTATTACATGGCGGAATAAANAATATTATGCTTC 4742
Qy 2635 gtggaggaaattataatgtgactaaatgtaaatattccacttgggaatcaggtgcgcacaa 2694
Db 4743 GATTAGGGATATATAATTTTATTCAACTATCGCTATGTTACTTGGGAAGCGGTGCGTCAAA 4802

Page 25

Search completed: September 14, 2000, 00:18:18
Job time: 3871 sec

THIS PAGE BLANK (USPTO)